STATE OF CALIFORNIA
DEPARTMENT OF MOTOR VEHICLES

EVALUATION OF THE CLASS C DRIVER LICENSE WRITTEN KNOWLEDGE TESTS

December 2006

Research and Development Branch
Licensing Operations Division

Authors: Thomas W. Reiner
Robert A. Hagge

RSS-06-221

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 2006	Final Report

4. TITLE AND SUBTITLE
Evaluation of the Class C Driver License Written Knowledge Tests

5. FUNDING NUMBERS

6. AUTHOR(S)
Thomas W. Reiner and Robert A. Hagge

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
California Department of Motor Vehicles
Research and Development Section
P.O. Box 932382
Sacramento, CA 94232-3820

8. PERFORMING ORGANIZATION REPORT NUMBER
CAL-DMV-RSS-06-221

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report represents the results of an evaluation of English and Spanish language written knowledge tests that were administered to applicants for an original or renewal Class C driver license. The report presents test fail rates for multiple attempts, mean error scores, and internal consistency validity for each test form, as well as the pass rate, item-choice selection rates, and item-total correlation for each item on each English test form. Items that need to be reviewed for possible rewording or replacement are identified. The results are based on 11,307 completed test forms collected from all California Department of Motor Vehicle field offices on August 4, 2005, or for a few offices on a subsequent Thursday.

14. SUBJECT TERMS
written test, fail rates, reliabilities, item statistics

15. NUMBER OF PAGES
54

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	None

PREFACE

This report is issued as an internal monograph of the Department of Motor Vehicles' Research and Development Branch rather than an official report of the State of California. The findings and opinions may not represent the views and policies of the State of California.

ACKNOWLEDGMENTS

This project was conducted under the general direction of David DeYoung, Research Chief, and the supervision of Robert Hagge, Research Manager II. Martha Boudreau, Manager III, provided specific information in relation to the written test content and format. Staff in the Abstracts Processing Section, under the supervision of Susie Wagner, key entered the data from the completed written tests. Debbie McKenzie, Associate Government Program Analyst, helped type and proofread the report.

SUMMARY

Introduction

- This report presents the results of an evaluation of the English DL 5 (Rev. 6/05), Spanish DL 5 (Rev. 6/05), and English DL 5T (Rev. 3/05) written knowledge examinations for a noncommercial Class C license. The study assessed the fail rate, mean number of errors, and internal-consistency reliability for each English or Spanish test form, as well as the pass rate, percentage of applicants selecting each answer choice, and item-total correlation for each item on each English test form. An assessment of the randomness of the correct answer choice assignment for the English DL 5 test forms is also presented.

- The results are based on 11,307 completed test forms that were collected from field offices in August 4, 2005, and for a few offices, on a subsequent Thursday.

Results

- The overall fail rates for first-attempt applicants are 50.1% for English originals, 33.6% for English renewals, 74.8% for Spanish originals, 79.8% for Spanish renewals, and 42.7% for English provisionals. The English fail rates are not much different from those obtained in the 2004 statewide evaluation, but the Spanish fail rates are higher than before, as shown in Figure 1.

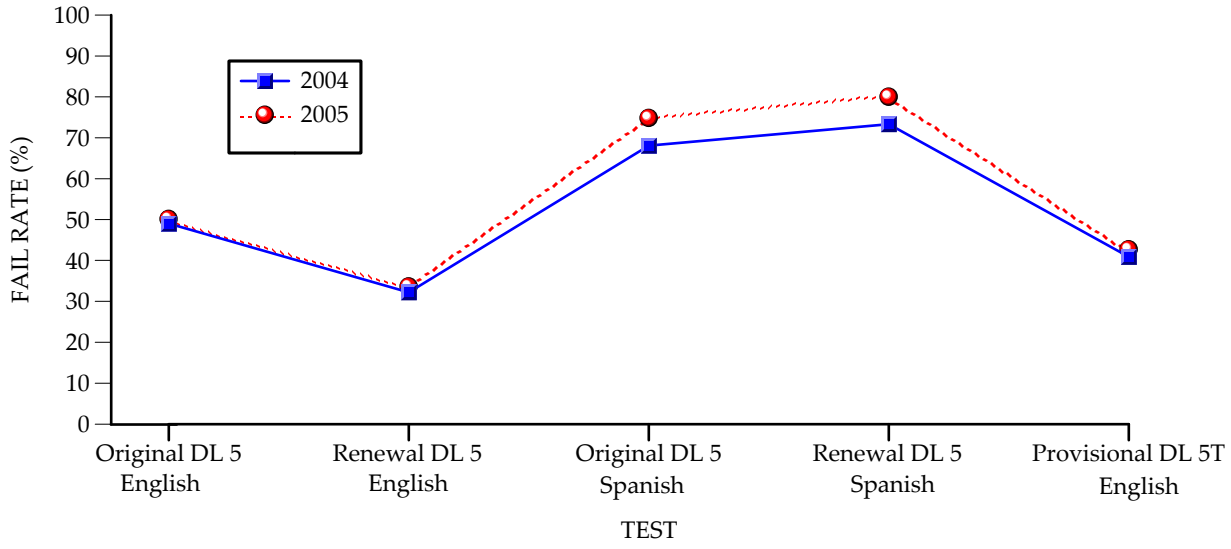

Figure 1. Comparison of first-attempt test fail rates for the 2004 and 2005 statewide evaluations.

- The fail rates for all test types tend to either remain about the same or increase over multiple attempts. The fail rates for the first through third attempts are shown in Figure 2.

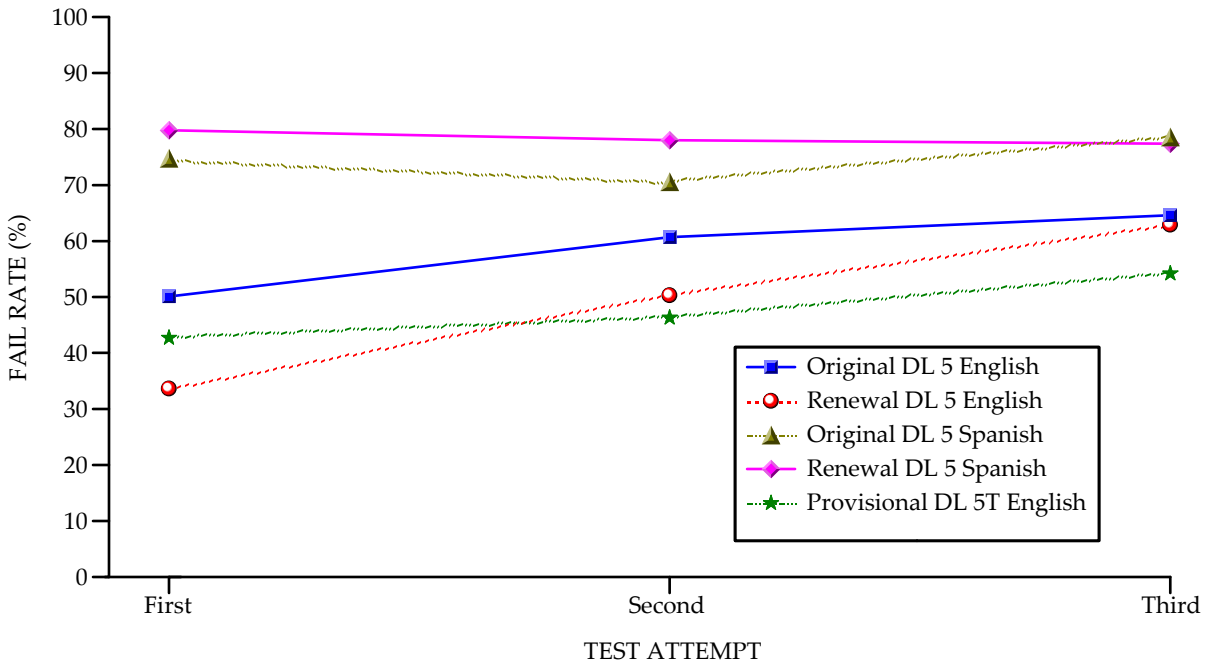

Figure 2. Fail rates for the first, second, and third test attempts on each test.

- The test forms vary considerably in difficulty, with form fail rates for any given test differing by as much as 34.8 percentage points.

- Nearly all of the internal-consistency test reliabilities for the English and Spanish DL5 forms for original applicants and the English DL 5T test forms for provisional applicants meet or exceed the .70 standard of acceptability. However, only two of the English renewal test forms and one Spanish renewal test form have reliabilities that meet this standard. The deficient renewal test reliabilities found in this and prior statewide evaluations are primarily due to the small number of items on these tests.

- Several questions on each test form are potentially deficient, as indicated by their having low item-total correlations, pass rates that are too high or too low, or distracter selection rates that are too high or too low.

- Answer choices a, b, and c were equally represented as correct answers on the original DL 5 English forms but not on the renewal DL 5 English forms.

- Some field offices administered the back of the DL 5 test sheet to renewal applicants, which is not consistent with the department's policy of using the front of the test form for renewals.

- Examiners changed some applicants' original answer choices when calculating the total test scores for original, renewal, and provisional applicants. This resulted in the examiner fail rates being considerably lower than the computer graded fail rates. Specifically, the fail rates for the English original, renewal, and provisional tests are 50.1%, 33.6%, and 42.7%, respectively, based on computer scoring, but only 45.8%, 24.7%, and 38.0% based on examiner scoring (for tests marked with an examiner score).

- Many field offices were using older test versions rather than the current revisions of the Class C license tests. This practice reduces the effectiveness of randomizing the order of items on the English DL 5 every 3 months to curtail applicant cheating.

- The readability level of the English and Spanish versions of the tests was found to be at or below the sixth-grade reading level. A reading level of fifth- or sixth-grade is considered optimum in communicating with driver license applicants who have difficulty reading, by standards of the American Association of Motor Vehicle Administrators.

- The percentage of first-attempt applicants who reported having read the California Driver Handbook ranges from 60.5% for Spanish renewals to 91.0% for English provisionals. The percentages were much higher for originals than for renewals.

- The amount of time applicants spent studying for each test ranges from 13.9 hours for Spanish renewals to 26.1 hours for Spanish originals. The mean hours of study is much higher for originals than for renewals.

Recommendations

- Test questions with characteristics that indicate they may be problematic should be reviewed and revised or replaced as necessary.

- Renewal applicants should have to complete all 36 questions on the DL 5 instead of only the first 18. This would bring the reliability of the renewal tests up to standard.

- The order of correct answer choices in each question should be randomized at least annually to decrease opportunity for cheating or rote memorization of the correct answers.

- Increase the reliability of the renewal test forms by doubling their length to 36 items.

- Take steps to ensure that field offices administer only the current version of each test.

- Reverse the policy that allows examiners to restate missed test questions when the applicant initially misses too many questions to pass.

- The department should continue investigating the possible use of computer technology to automate the creation and administration of the knowledge tests.

- A larger sample of the Spanish tests should be collected during the next statewide evaluation to make it possible to calculate item statistics for this language group.

- The department should make non-English versions of the driver handbook more readily available in both hardcopy and electronic form.

- The department should publicize, through press releases, information brochures, internet websites, and other means, the specific knowledge content areas that are most challenging to applicants and the importance of reading the driver handbook before taking the test.

- Procedures should be implemented to better identify applicants who have difficulty reading English and would be better served by being given an oral or non-English knowledge test.

- A waiting period (perhaps one week as currently required of applicants under age 18) should be required between test attempts to increase the likelihood that applicants would spend more time studying the driver handbook.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

APPENDICES

LIST OF TABLES

TABLE OF CONTENTS (continued)

LIST OF TABLES (continued)

LIST OF FIGURES

INTRODUCTION

This report presents the results of an evaluation of the English DL 5 (Rev. 6/05), Spanish DL 5 (Rev. 6/05), and English DL 5T (Rev. 3/05) written knowledge tests for a noncommercial Class C license. The findings will help guide the department in its ongoing efforts to maintain the validity of these tests.

There are 10 versions or forms of the English DL 5 and five forms of the Spanish DL 5. Each form has 36 questions. These tests are given to Class C applicants who are 18 years of age or older. Original driver license applicants complete all 36 questions, while driver license renewal applicants complete only the first 18. The DL 5T is given to original Class C license applicants who are younger than 18 years of age (provisional licensees). This test has five forms, each with 46 questions.

The entire item pool for the English DL 5 is comprised of 342 questions that were developed by subject matter experts within the department. These items cover only material contained in the California Driver Handbook and represent 23 different knowledge content areas. Items from each content area were assigned to the 10 forms of the English DL 5 in equal proportions. The number of items from each content area that are on each DL 5 form are shown in Table 1.

This report presents the fail rate, mean number of errors, and internal-consistency reliability for each form of the English and Spanish DL 5 tests and the English DL 5T test. It also includes the pass rate, percentage of applicants choosing each answer choice, and the item-total correlation for each item on each form of the English DL 5 and DL 5T tests. Item statistics are not presented for the Spanish DL 5 because too few test sheets of this type were collected to compute accurate estimates.

Table 1

Number of Items Assigned to the DL 5 Test Forms
from Each Knowledge Content Area

Content area	Total number of items	Number of items on each form
Accident responsibility	10	1
Driving in inclement weather	20	2
Driving on freeways	10	1
Driving under the influence	10	1
Driving with special vehicles	10	1
Improving traffic flow	10	1
Lane markings	10	1
Lane usage	10	1
Mandatory questions (BAC & vehicle sale)	2	2
Parking (general)	10	1
Parking on hills	10	1
Railroad crossings	10	1
Right-of-way	20	2
Road hazards	10	1
Safe driving practices	30	3
Seat belts and child restraints	10	1
Sharing the roadway with others	10	1
Space cushion (around vehicle)	20	2
Speed and speed limits	20	2
Traffic lights and signals	20	2
Traffic signs	30	3
Turns	20	2
Visual scanning	30	3
Total	342	36

Note. There are 10 equivalent forms of the English DL 5. Each form contains two-or-more items that relate to the interaction of vehicles and pedestrians. The items are typically drawn from either the right-of-way, safe driving practices, sharing the roadway with others, traffic lights and signals, or visual scanning content areas. The safe driving content area contains items relating to vehicle equipment usage (e.g., horn, headlights, turn signals, parking lights, and emergency flashers), general safe driving rules, accident avoidance and protection, defensive driving, driving when tired, and other subject matter.

METHODS

Data Collection

All DMV field offices were asked to send to the department's Research and Development Branch (R&D) all English and Spanish noncommercial Class C driver license written knowledge tests completed on Thursday, August 4th, 2005. A few offices failed to submit data for that day and were asked to collect tests on a subsequent Thursday. Tests were ultimately received from 170 of the 172 field offices that were open during the collection period. King City and Oxnard reported that they sent their completed tests to R&D, but they were never received.

The field offices were instructed to use the most current version of each test. The tests were processed in the usual manner, by marking incorrect responses, circling correct responses to missed items, and recording the total number of wrong answers. The office technicians were instructed to also write on the test sheet the test attempt number (1st, 2nd, 3rd, etc.) for the applicants' current original or renewal license processing, and "ORIG" or "REN" to indicate whether the applicant was applying for an original or renewal license. Field staff were also instructed to ask applicants the following two questions: "How long did you study for the test?" and "Did you read the DMV driver handbook?" Applicant responses to these questions were recorded on the test sheet.

Tests were received and screened by R&D. All information recorded on the test sheet and item responses were key-entered into an electronic file by staff from the department's Abstracts Processing Section. The tests were graded by computer to obtain accurate test and test-item statistics.

Data Analysis

A statistical technique known as analysis of variance (ANOVA) was used to determine whether any of the differences between test fail rates or average test scores are statistically significant. Differences are considered to be statistically significant if the probability (p) of their occurrence by chance alone is less than 5 times in 100. When significant differences were found from the omnibus ANOVA, Games and Howell

multiple-comparisons tests were subsequently used to determine what specific rates or means significantly differ from one another.

The internal-consistency reliability of each test form was computed using the Kuder-Richardson formula (KR-20). In general, this type of reliability indicates the degree of uniformity in subject matter content among test items, and commensurately the overall precision of the test as a measurement instrument. If a test has a high degree of reliability, a person should achieve a similar test score over repeated administrations of the test (assuming that the person's true knowledge level does not change between the tests.) The KR-20 reliability coefficient can range in value from 0 to 1. A value of 0 indicates that no similarity exists between the test items. A value of 1, on the other hand, denotes that the items on the test are perfectly homogeneous in content. Thus, coefficients closer to 1 indicate greater reliability and are more desirable. It is generally accepted by psychometricians that a reliability below .70 is too low and needs to be improved.

The item-total correlation coefficient is a measure of the degree that performance on an individual test item is related to performance on the entire test. The item-total correlation coefficient can range in value from -1 to 1. Items with a positive correlation coefficient value are more likely to be answered correctly by applicants with high test scores, whereas items with negative coefficient values are more likely to be answered correctly by applicants with low test scores. A coefficient value close to zero indicates that answering the test question correctly or incorrectly has very little or no relationship with whether an applicant scores high or low on the overall test, which may be due to wording ambiguity or some other problem with the question.

RESULTS

Data Collection and Screening

A total of 12,078 test forms of various types were received by R&D for the 1-day collection period. Some of these were excluded from the evaluation because they were old versions of the test (692), were renewal tests in which the back of the DL 5 was administered instead of the front (57), or their language, form number, or field office number were not key entered correctly (22). This left 11,307 usable test forms for the evaluation. It is very unlikely that the exclusion of the unusable forms biased the estimation of item and test statistics because the reasons why the forms were not usable do not seem to be related in any way to test performance.

It is necessary to have at least 100 first-attempt test sheets of a given test form to produce reasonably accurate estimates of item statistics for that form. This standard was met for the English DL 5 and DL 5T forms. However, it was not met for the Spanish DL 5 forms, so item statistics were not calculated for those.

All usable test forms were graded by computer, with an item being counted as incorrect if a wrong answer choice was marked, the item was left blank, or more than one answer choice was marked. The computer-graded scores were used to compute the test form fail rates and internal-consistency reliabilities. The fail rates are based on the current minimum passing standards that allow six errors for DL 5 original applicants, three errors for DL 5 renewal applicants, and eight errors for DL 5T provisional applicants.

Examiner Scoring Bias

Previous written test evaluations conducted by R&D have demonstrated that computer grading of the tests often produces results that differ from those based on the test score written by the examiner on the front of the test form. This occurs because departmental policy allows examiners to rephrase or restate missed questions and to award points based on the applicant's verbal responses, which often results in changing the outcome of the test to a pass. To determine the extent of the examiner scoring bias, three different fail rates were calculated. The first fail rate was calculated from computer grading of the tests. The second fail rate was calculated from the scores that the examiners wrote on the test forms, using only forms that had a score recorded. The third fail rate represents a blending of computer and examiner scoring, with the examiner score being used to determine the test outcome (pass or fail) when the score was written on the form, and the computer-graded score being used to determine the

outcome when the examiner score was not recorded. The three fail rates for the English DL 5 for originals and renewals, and the DL 5T for provisionals, are shown in Table 2 and illustrated in Figure 3.

Table 2

Number of Tests (*n*) and Fail Rates for the English Tests on the First Attempt When the Tests Were Graded by Computer, Examiner, and Examiner/Computer in Combination

Test type	Computer Graded		Examiner Graded		Examiner/ Computer Graded	
	n	Fail rate (%)	*n*	Fail rate (%)	*n*	Fail rate (%)
Original DL 5 English	2,138	50.1	1,983	45.8	2,138	45.5
Renewal DL 5 English	3,220	33.6	2,807	24.7	3,220	27.2
Provisional DL 5T English	1,238	42.7	1,160	38.0	1,238	38.1

Note. Examiner/computer grading used the examiner score if available and the computer score otherwise.

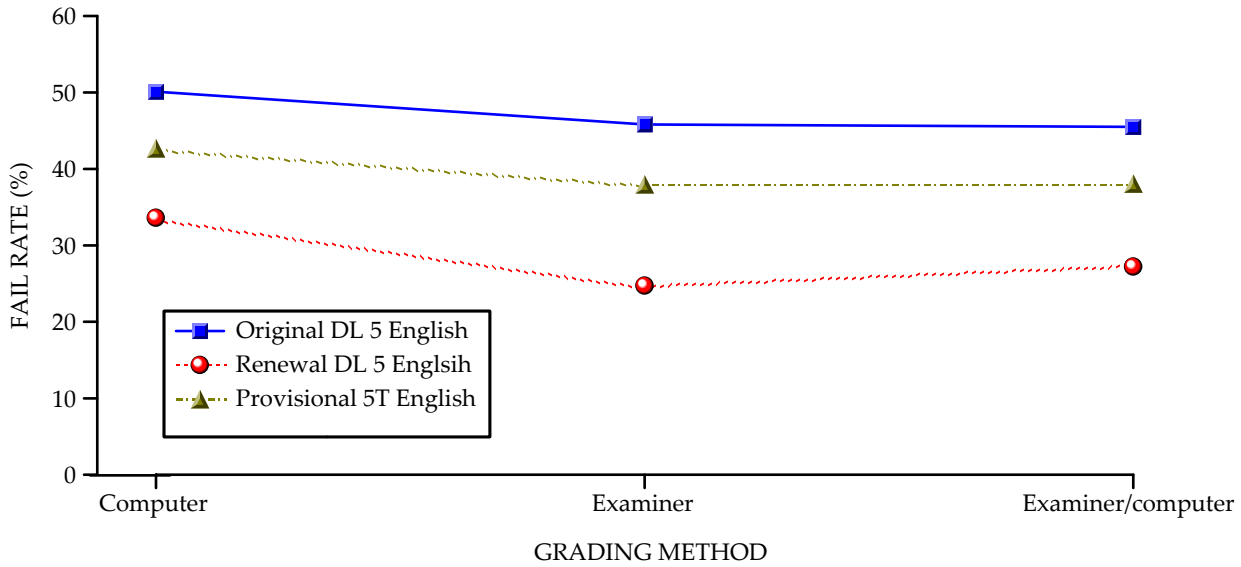

Figure 3. Fail rates for the English tests on the first test attempt for each method of grading.

The examiner-graded fail rates are lower than the computer-graded fail rates, indicating a scoring bias in favor of the applicant. This finding confirms that examiners have been exercising their authority to restate or rephrase missed questions, making it easier for applicants to pass.

Test Statistics

Test Form Difficulty and Reliability

The fail rate, mean number of errors, and internal-consistency reliability coefficient for each test form are presented in Table 3. The differences in the form fail rates and mean errors, and the pattern of internal-consistency reliabilities for the forms, are illustrated for each type of test in Figures 4, 5, and 6, respectively. The test statistics shown in the table and figures are discussed in the following five subsections.

One caveat that should be noted here is that some of the test forms that were collected would have been marked with an attempt number that is lower than what was true. This would have occurred if the field technician had recorded the test attempt on the current paid application rather than the attempt over all paid applications for original or renewal licensure. For example, a test marked as first attempt may actually have been an applicant's fourth try on the test—three of them on the first application and the fourth on the second application. Similarly, a test marked as second attempt may actually have been an applicant's fifth try—three on the first application and two on the second application. The extent to which this occurred and the net biasing influence this would have had on the estimation of fail rates is unknown.

Table 3

Number of Tests (*n*), Fail Rate, Mean Number of Errors, and Internal-Consistency Reliability Coefficient for Each Test Form on the First Test Attempt

Test form	*n*	Fail rate (%)	Mean errors	Reliability
Original DL 5 English[a]				
1	222	48.6	7.2	.80
2	225	48.9	7.1	.80
3	213	59.2	7.9	.78
4	217	48.8	7.4	.81
5	205	47.8	6.9	.76
6	200	53.5	8.1	.82
7	225	56.9	8.0	.82
8	200	52.0	7.6	.83
9	218	39.0	6.1	.79
10	213	46.9	6.9	.79
Total	2,138	50.1	7.4	.80
Renewal DL 5 English[b]				
1	311	24.1	2.4	.67
2	315	54.9	4.2	.56
3	322	37.0	3.3	.70
4	319	27.6	2.7	.71
5	339	27.4	2.6	.50
6	325	33.2	3.0	.56
7	321	47.0	3.7	.60
8	331	35.3	3.1	.66
9	313	20.1	2.2	.61
10	324	29.6	2.7	.63
Total	3,220	33.6	3.2	.62
Original DL 5 Spanish[c]				
1	66	68.2	9.4	.75
2	67	71.6	10.2	.80
3	65	80.0	11.1	.75
4	66	77.3	10.7	.80
5	53	77.4	12.4	.87
Total	317	74.8	10.8	.80
Renewal DL 5 Spanish[d]				
1	37	83.8	5.7	.32
2	40	70.0	5.4	.63
3	31	77.4	5.6	.53
4	27	74.1	5.4	.76
5	43	90.7	6.8	.65
Total	178	79.8	5.8	.58
Provisional DL 5T English[e]				
1	247	37.2	7.3	.74
2	254	46.1	8.5	.77
3	251	54.6	9.6	.80
4	247	37.7	7.4	.76
5	239	37.7	7.8	.81
Total	1,238	42.7	8.3	.78

Note. The figures presented for total fail rate, total mean errors, and total reliability are weighted averages. All ANOVAs are two-tailed.

[a]The forms differ significantly on fail rate ($F = 2.78$, $p < .01$). [b]The forms differ significantly on fail rate ($F = 16.63$, $p < .001$). [c]The forms do not differ significantly on fail rate ($F = 0.80$, $p = .53$). [d]The forms do not differ significantly on fail rate ($F = 1.66$, $p = .16$). [e]The forms differ significantly on fail rate ($F = 6.02$, $p < .001$).

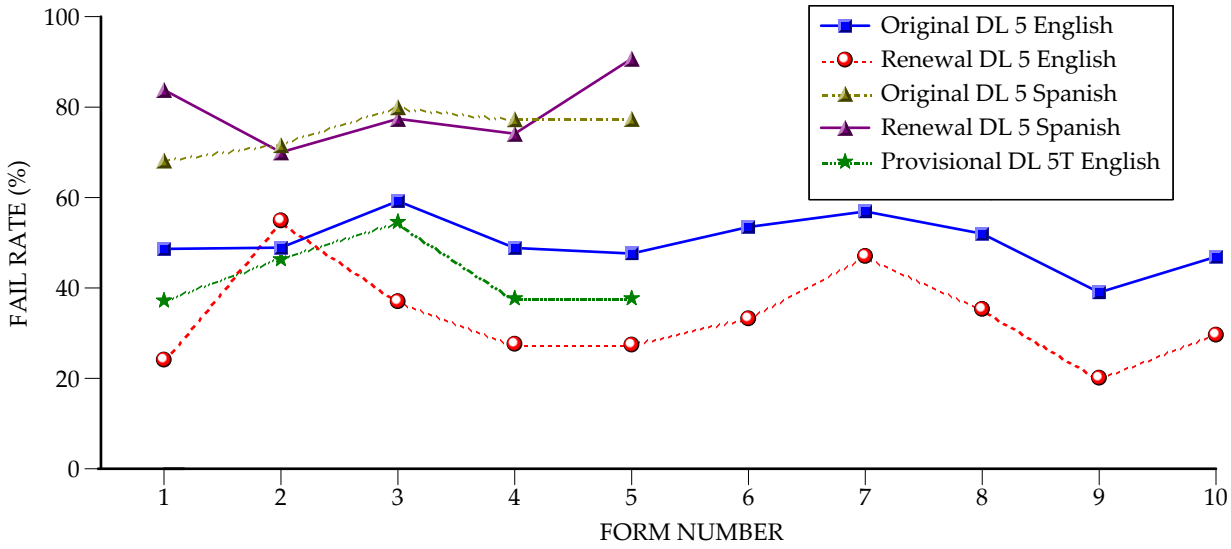

Figure 4. Test form fail rates for each test type on the first attempt.

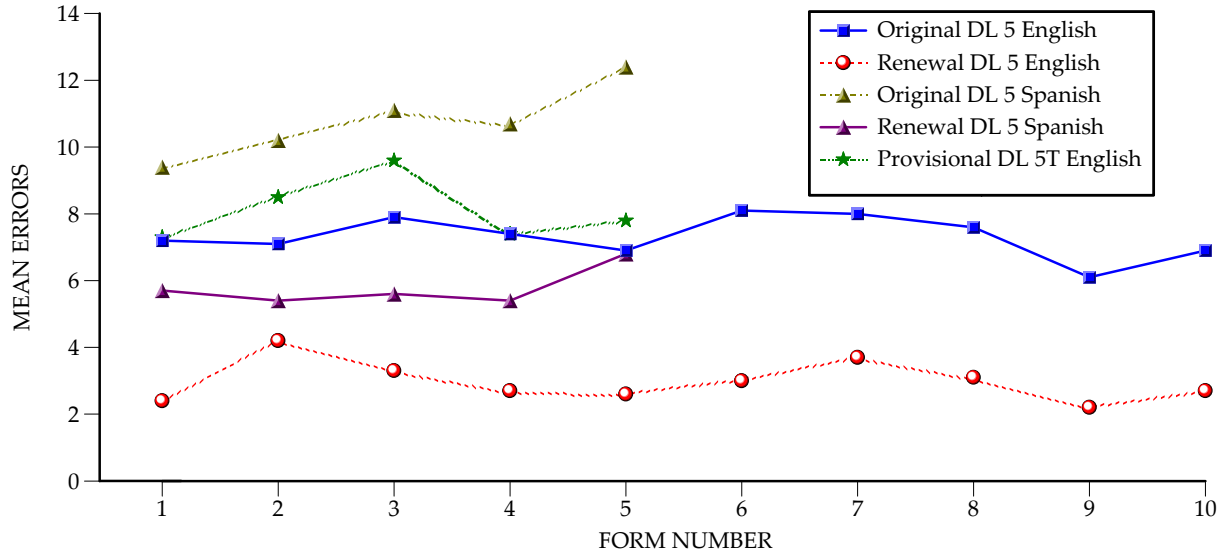

Figure 5. Mean numbers of errors on each form of each type of test on the first attempt.

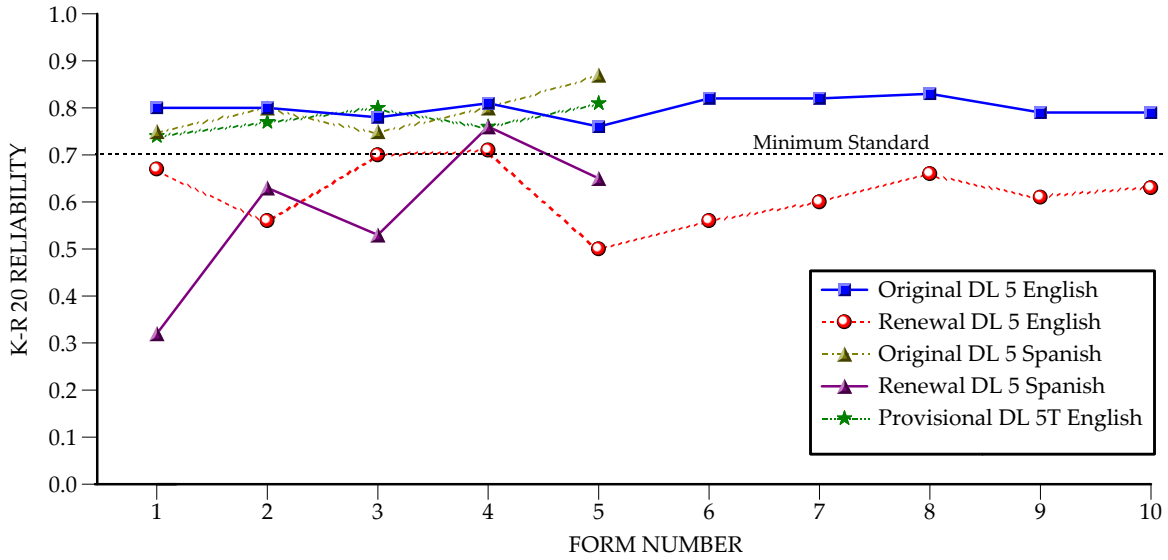

Figure 6. Internal-consistency reliability for each form of each test on the first attempt.

Original DL 5 English. The fail rates for the 10 forms of the English DL 5 taken by original applicants range from 39.0% to 59.2% and average 50.1% overall. The results of the omnibus ANOVA indicate that significant differences exist among the fail rates ($F = 2.78$, $p < .01$). The Games and Howell post hoc significance tests found that the fail rate for Form 9 is significantly lower than the fail rates for Forms 3 and 7. No other significant differences were found among the fail rates. The mean number of errors range from 6.1 to 8.1 and average 7.4 overall. (Statistical significance tests were not performed on test scores in the present evaluation.) All of the test reliabilities exceed the .70 whole-test reliability standard, with values ranging from .76 to .83.

Renewal DL 5 English. The fail rates for the 10 forms of the English DL 5 for renewal applicants range from 20.1% to 54.9% and average 33.6% overall. The differences between these rates are statistically significant overall ($F = 16.63$, $p < .001$). The Games and Howell tests found the following specific differences between the individual forms to be statistically significant: Form 2 higher than every other form except Form 7; Form 7 higher than every form except Forms 2, 3, and 8; and Form 9 lower than Forms 2, 3, 6, 7, and 8. The test score error means for the forms range from 2.2 to 4.2 and average 3.2 overall. The KR-20 reliabilities for the 10 forms range from .50 to .71, with all but two of

them falling below the .70 whole-test standard. These low coefficients suggest that 18 questions may not be enough to produce an adequate level of reliability, given that the difference in reliabilities between the 36-item test for original applicants and the 18-item test for renewals is primarily due to the difference in test lengths.

Original DL 5 Spanish. The form fail rates for original applicants taking the Spanish DL 5 range from 68.2% to 80.0% and average 74.8% overall. None of the rate differences are statistically significant ($F = 0.80$, $p = .53$). The test score error means range from 9.4 to 12.4 and average 10.8 overall. The reliability values for the five forms all exceed the .70 whole-test standard, with values ranging from .75 to .87.

Renewal DL 5 Spanish. The form fail rates for renewal applicants taking the Spanish DL 5 range from 70.0% to 90.7% and average 79.8% overall. None of the rate differences are statistically significant ($F = 1.66$, $p = .16$). The mean error scores range from 5.4 to 6.8 and average 5.8 overall. The KR-20 reliability values range from .32 to .76, with four of them falling below the .70 whole-test standard. These results once again reflect that the renewal test may be too short to have adequate reliability.

Provisional DL 5T English. The form fail rates for provisional applicants taking the English DL 5T range from 37.2% to 54.6% and average 42.7% overall. The differences between the rates are statistically significant ($F = 6.02$, $p < .001$). Specifically, the fail rate for Form 3 is significantly higher than the fail rates for Forms 1, 3, 4, and 5. The mean scores range from 7.3 to 9.6 and average 8.3 overall. All of the forms have a reliability level that exceeds the .70 standard, with the coefficients ranging from .74 to .81.

Test Difficulty by Attempt

The fail rate and mean number of errors for each test on each attempt are presented in Table 4. (Test sheets that do not have an attempt number recorded were excluded from the computation of these measures. This is unlikely to have biased the estimates because it is improbable that the non-reporting of attempt number is related in any way to test performance.) As has been found in prior statewide evaluations of the Class C written tests, the fail rate tends to remain steady or increase over subsequent test attempts. These results indicate that applicants very often either did not prepare for the retests by thoroughly reading the *California Driver Handbook*, or had problems reading and understanding these written materials.

Table 4

Number of Tests (*n*), Fail Rate, and Mean Number of Errors for Each Test Attempt

Test and attempt	*n*	Fail rate (%)	Mean errors
Original DL 5 English			
1st	2,138	50.1	7.3
2nd	898	60.7	8.2
3rd	412	64.6	8.7
4th or higher	406	55.2	7.9
Not reported	239	46.9	7.1
Renewal DL 5 English			
1st	3,220	33.6	3.0
2nd	581	50.3	4.0
3rd	132	62.9	4.9
4th or higher	332	35.5	3.3
Not reported	298	30.5	3.0
Original DL 5 Spanish			
1st	317	74.8	10.7
2nd	231	70.6	9.6
3rd	103	78.6	10.4
4th or higher	87	60.9	8.8
Not reported	50	66.0	9.2
Renewal DL 5 Spanish			
1st	178	79.8	5.8
2nd	123	78.1	5.8
3rd	53	77.4	5.9
4th or higher	42	83.3	6.4
Not reported	23	86.9	6.5
Provisional DL 5T English			
1st	1,238	42.7	8.1
2nd	447	46.3	8.7
3rd	166	54.2	9.3
4th or higher	189	39.1	7.9
Not reported	145	41.3	8.3

Note. All figures presented for total fail rate and total mean errors are weighted averages.

Test Fail Rates by Field Office

The number of English original and renewal tests received and the fail rates for these tests across all test attempts are presented for each field office in Appendix A. (Field office fail rates computed from fewer than 20 test forms tend to lack precision and therefore may not be trustworthy.) Fail rates for the Spanish DL 5 tests are not presented because too few of these tests were collected to compute accurate estimates. The English test fail rates for field offices with 20 or more forms range from 19.5% to 86.7% for original applicants and from 11.1% to 69.2% for renewal applicants. These

rates indicate wide variation in the knowledge level of applicants residing in different geographical areas of the state.

Assessment of Answer Choice Randomness

The number and percentage of times each answer choice (a, b, or c) is the correct answer on each English test form are presented in Table 5. If the representation of each answer choice as being correct is truly random, each answer choice should be the correct answer 33% of the time across all items on each individual test form and also across all test forms combined.

Table 5

Number (*n*) and Percentage of Times that Each Answer Choice was the Correct Answer on Each Form of the Original and Renewal DL 5 English Tests

Test type and form	Number of questions	Answer choice					
		a		b		c	
		n	%	*n*	%	*n*	%
Original[a]							
1	36	13	36.1	12	33.3	11	30.6
2	36	16	44.4	9	25.0	11	30.6
3	36	9	25.0	14	38.9	13	36.1
4	36	12	33.3	17	47.2	7	19.4
5	36	9	25.0	11	30.6	16	44.4
6	36	10	27.8	10	27.8	16	44.4
7	36	10	27.8	14	38.9	12	33.3
8	36	13	36.1	13	36.1	10	27.8
9	36	12	33.3	17	47.2	7	19.4
10	36	12	33.3	15	41.7	9	25.0
Total	360	116	32.2	132	36.7	112	31.1
Renewal[b]							
1	18	5	27.8	8	44.4	5	27.8
2	18	8	44.4	5	27.8	5	27.8
3	18	6	33.3	6	33.3	6	33.3
4	18	6	33.3	10	55.6	2	11.1
5	18	3	16.7	7	38.9	8	44.4
6	18	6	33.3	4	22.2	8	44.4
7	18	5	27.8	9	50.0	4	22.2
8	18	4	22.2	9	50.0	5	27.8
9	18	6	33.3	12	66.7	0	0.0
10	18	6	33.3	8	44.4	4	22.2
Total	180	55	30.6	78	43.3	47	26.1

Note. All χ^2 tests are two-tailed.

[a]The three answer choices do not differ significantly in how often they are the correct answer across all 10 test forms combined ($\chi^2[2, N = 360] = 1.87$, $p = .39$). The proportional distribution of the answer choices being correct on each separate test form does not differ significantly between the 10 test forms ($\chi^2[18, N = 360] = 18.67$, $p = .41$). [b]The three choices differ significantly in how often they are the correct answer choice across all 10 test forms combined ($\chi^2[2, N = 180] = 8.63$, $p < .05$). The proportional distribution of the answer choices being correct on each separate test form differs significantly between the 10 test forms ($\chi^2[18, N = 180] = 29.00$, $p < .05$).

For the English DL 5 taken by originals, the differences in the proportional representations of the three answer choices being correct over all 10 test forms combined are not statistically significant ($p = .39$). In addition, the percentage distribution of the answer choices being correct on each separate form does not significantly differ between the 10 forms ($p = .41$). These findings indicate that the answer choices are adequately randomized on these forms.

For the English DL 5 taken by renewals, the differences in the proportional representations of the three answer choices being correct over all 10 test forms combined are statistically significant ($p < .05$), and the percentage distribution of the answer choices being correct on each separate form differs significantly between the 10 forms ($p < .05$). These findings indicate that the correct answer choices are not completely randomized on these test forms.

Item Statistics

All item statistics presented in this report are based on first-attempt tests. The results of the item analyses for the English original, renewal, and provisional tests are presented in Appendices B, C, and D, respectively. As indicated earlier in this report, item statistics are not presented for the Spanish tests because not enough of these were received for accurate estimation. Each of these appendices has four tables that contain, respectively, item pass rates and answer choice selection rates; item-total correlations; the percentage of applicants who would pass at different cut-points; and a listing of problem items identified on each test form. The results presented in these tables are described in the following four subsections of this report.

Item Pass Rates and Answer Choice Selection Rates

The answer choice selection rates are presented in Tables B1, C1, and D1. The percentages for the correct answer choices are underlined in the tables. These rates are based on valid responses only. Not represented are instances in which the applicant did not answer the item or selected more than one answer choice. These latter cases represent, for any given item, fewer than 3% of the applicants taking the test.

The item pass rate refers to the percentage of applicants who correctly answered the question. A very low item pass rate may indicate that the item is poorly worded, has

ambiguous or misleading answer choices, is not related to the general knowledge domain being tested, or is problematic for some other reason. Items in which 60% or fewer of the applicants answered correctly are generally considered suspect and should be reviewed for clarity and accuracy. Items passed by 95% or more of the applicants are also questionable and may need to be revised, because this often indicates that the distracters are so illogical that the correct answer is obvious or that the subject matter is common knowledge and therefore doesn't need to be tested. The items with pass rates that are too high or too low are shaded in the tables.

It is also desirable that the incorrect answer choices seem tenable to applicants lacking the knowledge covered by the item. Therefore, items with individual distracters selected by 2% or fewer of the applicants should also be considered for possible revision. Distracters selected more often than the correct answer, or within 10% of the selection rate for the correct answer, are also suspect and should be reviewed. Response rates that are too low or too high are shown in boldface type in the table.

The 60% and 95% pass rate criteria for items, and the 2% selection rate criterion for distracters, are provided only as guidelines. It is not recommended that items be revised or replaced on the basis of these criteria alone. The relative importance of knowledge covered by an item, possible wording problems, and other relevant factors should always be weighed when deciding to revise or replace an item. For example, almost all applicants may have knowledge of certain laws and principles covered by a test question. However, if such knowledge is considered critical to safe driving, the item should not be automatically discarded simply because more than 95% of the population correctly answers the item.

Item-Total Correlations

The item-total correlation coefficients for the test items are presented in Tables B2, C2, and D2. Items that tended to be answered correctly by applicants who scored low on the test overall (i.e., with a negative coefficient value), or that have very little or no relationship to the other items on the test (i.e., with a positive coefficient value less than .10), are highly suspect and almost always need to be modified or replaced. These items are shaded in the tables.

Percentage of Applicants Who Would Pass at Different Cut-Points

The percentages of applicants who would pass the tests at different score cut-points are presented in Tables B3, C3, and D3. The tables present the percentage of applicants who missed at least the number of items indicated in the leftmost column of each row, and would therefore have passed if that number had been used as the minimum passing score. For instance, Table B3 indicates that 40.5% of originals who completed Form 1 of the English DL 5 would have passed if the number of allowable errors had been five instead of six. The shaded row in each table shows the pass rate for each test form using the current passing standard (six for DL 5 original applicants, three for DL 5 renewal applicants, and eight for DL 5T provisional applicants).

Summary of Problem Items on Each Test Form

Tables B4, C4, and D4 identify the items with low item-total correlations, low or high pass rates, and distracters with low or high selection rates. As can be seen, all of the test forms have several items with one or more of these characteristics and should therefore be reviewed to determine if they should be revised or replaced.

Studying for the Test

Applicants Who Read the Driver Handbook

Table 6 presents the percentage of first-attempt applicants in each applicant group who did not provide a usable ("yes" or "no") response to the question of whether they read the driver handbook, the number of usable responses, and the percentage of applicants providing usable responses who said they studied the handbook.

The percentage of applicants who did not provide a usable response is less than 10% for all groups. Because these non-response rates are small and do not vary by much between applicant groups, any bias resulting from excluding these cases is believed to be small.

Table 6

Percentage of Applicants Who Did Not Answer "Yes" or "No" as to Whether They Read the Driver Handbook, the Number of Usable Responses, and the Percentage of Applicants Who Gave a Usable Response Who Said They Read the Driver Handbook, for Each Group of Applicants Taking a First-Attempt Test

Applicant group	% who did not answer "yes" or "no"	Number of usable responses	Of usable responses, % who read the driver handbook
English originals	7.5	1,977	80.6
English renewals	8.3	2,952	64.7
Spanish originals	8.5	290	87.6
Spanish renewals	8.9	162	60.5
English provisionals	6.5	1,157	91.0

Note. Applicants who did not answer "yes" or "no" as to whether they read the handbook were excluded from the computation of reading rates. The applicant groups differ significantly in how often they reported having studied the handbook ($F = 106.82$, $p < .001$).

The percentage of applicants who said they read the handbook ranges from 60.5% for Spanish renewals to 91.0% for English provisionals. The differences between these percentages are statistically significant ($F = 106.82$, $p < .001$). The results of post hoc Games and Howell significance tests indicate that English provisionals studied the handbook at a significantly higher rate than every other group except Spanish originals; Spanish originals studied at a significantly higher rate than every group except English provisionals; and English originals studied at a significantly higher rate than English and Spanish renewals. No other between-group differences are statistically significant.

Table 7 presents the same measures, comparing those who passed the test to those who failed the test in each applicant group.

Table 7

Percentage of Applicants Who Did Not Answer "Yes" or "No" as to Whether They Read the Driver Handbook, the Number of Usable Responses, and the Percentage of Applicants Who Said They Read the Driver Handbook, for Test Passes and Fails in Each Applicant Group on the First Test Attempt

Applicant group and test result	% who did not answer "yes" or "no"	Number of usable responses	Of usable responses, % who read the driver handbook
English originals			
Pass	6.3	999	80.7[a]
Fail	8.8	978	80.5
English renewals			
Pass	8.0	1967	67.2[b]
Fail	9.0	985	59.7
Spanish originals			
Pass	10.0	72	94.5[c]
Fail	8.0	218	85.3
Spanish renewals			
Pass	8.3	33	69.7[d]
Fail	9.2	129	58.1
English provisionals			
Pass	6.6	662	93.1[e]
Fail	6.4	495	88.3

Note. Applicants who did not answer "yes" or "no" were excluded from the computation of the percentage who said they read the driver handbook.

[a]The difference between the percentages is not statistically significant (χ^2 [1, N = 1,977] = .01, p = .91). [b]The difference is statistically significant (χ^2 [1, N = 2,952] = 16.22, p < .001). [c]The difference is statistically significant (χ^2 [1, N = 290] = 4.14, p < 05). [d]The difference is not statistically significant (χ^2 [1, N = 162] = 1.47, p = .23). [e]The difference is statistically significant (χ^2[1, N = 1,157] = 7.87, p < .01).

The difference between passes and fails in their likelihoods of reading the handbook is statistically significant for only English renewals (p < .001), Spanish originals (p < .05), and English provisionals (p < .01). (Small sample sizes reduced the likelihood of finding a statistically significant difference for Spanish renewals.) These results indicate that

applicants in these three groups tended to perform better on the test if they had studied for it. Only a small percentage of applicants did not answer "yes" or "no" as to whether they read the handbook, and there is little difference between test passes and fails in each group on this measure. Therefore, any bias in the estimation of the percentages who read the handbook resulting from excluding non-response cases is likely to be very small.

Hours of Study

Applicants were asked: "How long did you study for the test?" Since the focus of this question is not limited to study of the driver handbook, the answers reflect how long the applicants studied using all sources available to them. Table 8 shows, for first-attempt tests, the percentages of applicants in each applicant group whose responses were excluded from the estimation of mean hours studied because they gave an unusable text response, said they studied over 100 hours (too high to be trustworthy), or did not respond. These cases were excluded from the computation of mean hours of study. The number of usable responses and the mean number of hours that each group studied are also shown in the table.

Table 8

Percentage of Applicants Who Gave an Unusable Text Response, Who Said They Studied Over 100 Hours, Who Did Not Respond, the Number of Usable Responses, and the Mean Number of Hours of Study, for Each Applicant Group on the First Test Attempt

Applicant group	% who gave an unusable text response	% who said they studied over 100 hours	% who did not respond	Number of usable responses	Mean hours of study
English originals	24.2	5.4	2.0	1,462	23.0
English renewals	25.2	3.3	2.4	2,225	17.1
Spanish originals	26.8	9.2	1.6	198	29.8
Spanish renewals	22.5	6.2	4.0	120	13.9
English provisionals	22.4	10.9	1.3	810	27.0

Note. Mean hours of study are based on usable responses only.

The applicant groups differ significantly in the average amount of time they spent studying. The group means on hours of study range from 13.9 for Spanish renewals to 29.8 for Spanish originals. The differences in study times are statistically significant overall ($F = 43.24$, $p < .001$). Results of post hoc Games and Howell statistical significance tests show that Spanish originals studied longer than every other group except English provisionals, English provisionals studied longer than every other group except Spanish originals, and English originals studied longer than English and Spanish renewals. No other differences between applicant groups are statistically significant.

Since the groups have about the same percentage of applicants who gave an unusable text response, any bias in the computation of mean hours of study resulting from the exclusion of these specific cases is likely to be very small.

The groups vary more widely in how often respondents said they studied more than 100 hours. The amount and direction of any bias in the estimation of study times that may have resulted from excluding these cases is unknown. However, it is plausible that some of the applicants who said they studied over 100 hours actually did study that long (or at least more than average). If this is true, then excluding responses over 100 hours probably resulted in an underestimation of the true mean number of hours studied, and adjusting the means upward to counteract this bias would result in a greater spread between the means than what is shown in the table. However, since the true study time for applicants who reported over 100 hours is unknown, it is uncertain how large of a bias, if any, resulted from excluding these cases.

The percentage of applicants for whom responses were not recorded, either because they chose not to respond or the technician administering the test forgot to ask the question or write the answer on the test form, was very small for each applicant group. Therefore, any bias in the mean hours of study between the groups resulting from the exclusion of these cases is considered to be trivial.

Table 9 presents the same variables for test passes and fails separately. The difference in the amount of time passes and fails spent studying before taking the test is statistically significant only for English renewals ($p < .01$) and Spanish originals ($p < .05$).

Table 9

Percentage of Applicants Who Gave an Unusable Text Response, Who Said They Studied Over 100 Hours, Who Did Not Respond, the Number of Usable Responses, and the Mean Number of Hours of Study, for Passes and Fails in Each First-Attempt Applicant Group

Applicant group and test result	% who gave an unusable text response	% who said they studied over 100 hours	% who did not respond	Number of usable responses	Mean hours of study
English originals					
Pass	21.8	5.2	2.2	756	22.6[a]
Fail	26.7	5.6	1.9	706	23.5
English renewals					
Pass	24.1	2.7	2.2	1518	18.0[b]
Fail	27.2	4.5	2.9	707	15.2
Spanish originals					
Pass	23.8	15.0	1.3	48	37.2[c]
Fail	27.9	7.2	1.7	150	27.4
Spanish renewals					
Pass	13.9	8.3	2.8	27	18.3[d]
Fail	24.7	5.6	4.2	93	12.7
English provisionals					
Pass	21.4	11.2	1.4	468	27.6[e]
Fail	23.6	10.6	1.1	342	26.1

Note. Mean hours of study are based on usable responses only.

[a]The two group means are not significantly different (t[1,460] = -.76, p = .45). [b]The means are significantly different (t[2,223] = 2.79, p < .01). [c]The means are significantly different (t[196] = 2.19, p < . 05). [d]The means are not significantly different (t[118] = 1.62, p = .11). [e]The means are not significantly different (t[808] = .90, p = .37).

There is the potential for bias in the estimation of study times. The percentage of applicants who gave an unusable text response differs greatly between passes and fails for Spanish renewals. The extent of bias, if any, that this would have created is unknown. The passes and fails in each of the other applicant groups did not differ by much in how often they gave unusable text responses, so any bias this created in the comparisons for these applicant groups is considered to be small.

The percentage of applicants who reported studying more than 100 hours is relatively small for English originals and renewals and Spanish renewals, so any bias in the mean hours estimates caused by the exclusion of these cases is considered small for these groups. The percentages of cases deleted for this reason are larger for Spanish originals and English provisionals, and therefore there would be a greater downward bias in the average study times for passes and fails in these two groups than in the other three groups.

The difference between passes and fails in the percentage of applicants who estimated studying more than 100 hours is relatively large for Spanish originals, which may have resulted in the difference in the average study times for passes and fails being smaller than it really was (again assuming that those that reported more than 100 hours tended to study longer on average than those who did not).

DISCUSSION

The results of this evaluation show that the first-attempt fail rates for the English tests did not change much from what they were in the 2004 evaluation, but the fail rates for the Spanish tests are higher than they were before. The much higher fail rates for the Spanish tests has been a continuing concern to the department. While several possible explanations for this disparity can be hypothesized, no research has been conducted to identify the specific causes. The Spanish tests have the same subject matter content as the English tests, as they are translated directly from the English test forms. The department conducted a thorough review of the Spanish tests a few years ago to determine whether faulty translation could possibly explain the high Spanish test fail rates, and no significant problems were found. Lack of study can also be ruled out as an explanation for the higher Spanish test fail rates based on the finding in this evaluation that applicants who took the Spanish tests reported having studied the driver handbook as much as those who took the English tests. Other explanations will not be offered here because doing so without the supporting research would be wholly speculative.

The results also show no improvement in the reliability of the renewal test forms since the 2004 evaluation; all but 2 of the 10 forms fall below the .70 standard of acceptability. This was expected because the small number of questions on these tests makes them prone to be unreliable. This is a significant deficiency because it reduces the department's ability to make valid licensing decisions.

The overall randomness of answer choices a, b, and c being correct could be improved on the DL 5 renewal test. Increasing such randomness would increase the psychometric integrity of the tests because it would reduce the possibility of cheating or trying to pass the test by memorizing the pattern of correct answers. The current practice of randomizing the order of questions on the test each year also helps in this regard and should be continued.

The analysis of individual test questions found that all test forms contained some items with low item-total correlations, pass rates that were too high or too low, and distracters that were selected too often or too infrequently. The overall quality of the tests can be increased by reviewing these items and then revising or replacing them as

necessary. Items with weak item-total correlations are especially suspect and warrant immediate attention.

The results also show that original applicants tended to report studying the handbook more often and for longer periods than did renewal applicants. It was also found that those who passed the test generally tended to more often have read the driver handbook and studied longer than those who failed. This suggests an opportunity for the department to reduce fail rates by advising license applicants to thoroughly read the DMV handbook and spend more time studying for the test. Of course, the department must then make sure that the driver handbook is readily available, in different languages, for this to be effective.

Some inconsistent practices followed by field office personnel were discovered over the course of this evaluation and should be discontinued. During the data screening process, it became apparent that some field office personnel were administering the back side rather than the front of the DL 5 test sheet to renewal applicants. In addition, some field offices also administered older revisions of the tests. This practice diminishes the effectiveness of randomizing the English DL 5 test every 3 months to curtail cheating. Administering the correct side of the renewal test and using only current tests would help increase the overall integrity of the knowledge tests.

Another action that could be taken to improve the integrity of the testing process would be to stop the practice of restating or substituting questions when an applicant initially misses too many questions to pass. This practice results in licensing more applicants with marginal knowledge competency. Evidence of this is provided by the substantial increase in pass rates that was found when examiner error scores rather than computer-graded error scores were used to determine test outcomes in this study. This practice undermines the department's ongoing efforts to maximize the reliability and validity of the tests through periodic evaluations and the use of state-of-the-art psychometric test construction techniques.

RECOMMENDATIONS

There are several things the department can do to improve the quality of the written knowledge tests. The following specific actions are recommended.

1. Questions with one-or-more of the following statistical characteristics should be reviewed and revised or replaced as necessary: (a) an item-total correlation less than .10, (b) a distracter selected more often or within 10% of the correct answer, or by fewer than 2% of the respondents, or (c) a pass rate that is too high (95% or higher) or too low (60% or lower).

2. The order of answer choices should be periodically randomized to decrease the opportunity for cheating or rote memorization of the correct answers. Computer applications are available that can do this efficiently and cost-effectively. The department's Research and Development Branch can provide guidance in this matter if requested.

3. The renewal test form reliabilities should be increased to at least .70. This can be easily accomplished by having renewal applicants complete all 36 items on the DL 5 instead of only the first 18 items. Correcting problem items, especially those with low item-total correlations, would also increase test reliability but probably not enough.

4. Steps should be taken to ensure that field office personnel are administering only the current version of each test in accordance with the procedures stated in the department's Driver License Manual.

5. The department should reverse the policy that allows examiners to restate missed test questions when the applicant initially misses too many questions to pass. This would increase the integrity of the testing process by eliminating examiner subjectivity in determining whether the customer has an adequate understanding of traffic laws and safe driving practices. It would also protect the department against charges of discrimination based on the gender, age, race, or other characteristics of the applicant rather than their level of knowledge. R&D will provide a follow-up issue paper presenting further rationale for this recommendation upon request.

6. The department should continue investigating the possible use of computer technology to generate the test forms from a large item-pool database and possibly create a unique test for each applicant. Such technology would also support ongoing randomization of the order of test questions and answer choices.

7. A larger sample of Spanish language tests should be collected in the next statewide evaluation to make it possible to calculate item statistics for this language group. Although the Spanish tests are intended to be "exact" translations of the English versions, it is possible that good questions can become problematic upon translation. Such items could be identified by performing the standard item analyses.

8. The department should take steps to make the Spanish version of the driver handbook more readily available both in hardcopy and electronic form. This might include better publicizing how the handbook material can be obtained. (This should also be done to the extent possible for the other language translations of the handbook.)

9. The department should prepare and distribute information materials that emphasize the importance of thoroughly reading the driver handbook before taking the test. The specific knowledge content areas that are most challenging to applicants should also be publicized on the internet, in newspapers, and on television.

10. Procedures should be developed and implemented to better identify applicants who have difficulty reading English and who would benefit by taking the test orally or in another language. Of course, the most direct way would be to just ask applicants whether they would prefer to take an oral or non-English test. However, for this simple approach to be effective, applicants should first be made aware that they will not be penalized just because they are not able to read in English. The department should implement the means necessary to publicize this fact.

11. Applicants who fail a test should be required to wait a minimum period of time before retesting. Current law requires provisional license applicants to wait at least a week between written tests (and 2 weeks between drive tests). At a minimum, applicants should not be allowed to take more than one knowledge test per day. Instituting a mandatory waiting period would increase the likelihood that applicants would study the driver handbook more thoroughly before taking the test again.

APPENDICES

Appendix A

Test Fail Rates Over All Attempts on the English DL5 for Original and Renewal Applicants in Each Field Office

Table A

Number of Tests (*n*) and Test Fail Rates Over All Attempts on the English DL 5 (Rev. 6/05) for Original and Renewal Applicants in Each Field Office

Reporting unit number and field office	Originals		Renewals	
	n	Fail rate (%)	*n*	Fail rate (%)
501 Sacramento	36	38.9	31	29.0
502 Los Angeles	87	57.5	19	47.4
503 San Francisco	77	19.5	48	18.8
504 Oakland	54	31.5	45	28.9
505 Fresno	18	50.0	9	55.6
506 San Diego	67	41.8	49	34.7
507 Long Beach	53	62.3	60	36.7
508 Hollywood	106	50.0	48	45.8
509 Pasadena	58	53.4	60	33.3
510 Glendale	51	39.2	54	31.5
511 Montebello	48	56.3	39	46.2
512 San Bernardino	30	86.7	30	43.3
513 Truckee	3	0.0	7	28.6
514 Culver City	64	48.4	70	42.9
515 Van Nuys	49	61.2	69	39.1
516 San Jose	57	57.9	36	36.1
517 Stockton	37	64.9	31	48.4
518 Mountain View	0	*	0	*
519 San Diego-Clairemont	81	39.5	56	26.8
520 Chico	4	75.0	19	31.6
521 Jackson	2	0.0	13	30.8
522 Oroville	11	54.5	13	23.1
523 Concord	9	55.6	33	30.3
524 Crescent City	3	33.3	6	16.7
525 Placerville	5	40.0	20	20.0
526 Eureka	7	85.7	15	26.7
527 El Centro	17	76.5	7	28.6
528 Blythe	0	*	0	*
529 Bakersfield	25	68.0	19	57.9
530 Lakeport	3	33.3	11	72.7
531 Susanville	2	50.0	14	50.0

Table A (continued)

Number of Tests (*n*) and Test Fail Rates Over All Attempts on the English DL 5 (Rev. 6/05) for Original and Renewal Applicants in Each Field Office

Reporting unit number and field office	Originals		Renewals	
	n	Fail rate (%)	*n*	Fail rate (%)
532 Pomona	37	56.8	51	49.0
533 Madera	2	50.0	22	54.5
534 Corte Madera	11	45.5	32	21.9
535 Ukiah	12	58.3	7	85.7
536 Merced	5	40.0	35	57.1
537 Alturas	0	*	0	*
538 South Lake Tahoe	2	50.0	9	44.4
539 Salinas	19	57.9	16	50.0
540 Napa	7	71.4	16	50.0
541 Grass Valley	5	40.0	17	41.2
542 Santa Ana	67	62.7	45	37.8
543 Roseville	19	42.1	41	29.3
544 Quincy	0	*	4	25.0
545 Riverside	20	65.0	38	31.6
546 Hollister	4	75.0	5	0.0
547 San Luis Obispo	2	50.0	18	50.0
548 Redwood City	36	38.9	34	29.4
549 Santa Barbara	10	30.0	15	46.7
550 Capitola	13	15.4	19	47.4
551 Redding	11	36.4	35	31.4
552 Yreka	2	0.0	5	0.0
553 Tulelake	0	*	3	0.0
554 Vallejo	25	60.0	19	52.6
555 Santa Rosa	13	46.2	46	34.8
556 El Cerrito	40	57.5	52	44.2
557 Modesto	12	58.3	35	45.7
558 Red Bluff	4	25.0	17	41.2
559 Visalia	10	40.0	13	15.4
560 Ventura	14	28.6	34	20.6
561 Woodland	8	62.5	11	36.4
562 Yuba City	22	54.5	30	16.7

Table A (continued)

Number of Tests (*n*) and Test Fail Rates Over All Attempts on the English DL 5 (Rev. 6/05) for Original and Renewal Applicants in Each Field Office

Reporting unit number and field office	Originals		Renewals	
	n	Fail rate (%)	*n*	Fail rate (%)
563 Santa Maria	20	55.0	22	31.8
564 Colusa	0	*	0	*
565 Hanford	25	64.0	20	20.0
566 Mariposa	1	100.0	8	50.0
567 Seaside	24	58.3	38	26.3
568 San Andreas	4	25.0	14	28.6
569 Sonora	3	0.0	13	30.8
570 Auburn	2	0.0	26	15.4
571 Willows	1	0.0	1	0.0
572 Weaverville	1	0.0	2	0.0
573 Porterville	13	69.2	11	54.5
574 Paso Robles	1	100.0	15	46.7
575 Taft	0	*	15	33.3
576 Bell Gardens	74	63.5	27	55.6
577 Ridgecrest	4	50.0	7	14.3
578 Indio	34	70.6	20	30.0
579 Hayward	45	66.7	43	34.9
580 Clovis	17	58.8	36	36.1
581 Compton	12	100.0	9	100.0
582 Barstow	7	71.4	17	47.1
583 Watsonville	6	83.3	4	25.0
584 Needles	1	100.0	1	100.0
585 Bishop	1	0.0	0	*
586 Norco	35	62.9	51	25.5
587 Arleta	64	59.4	21	71.4
588 Vacaville	13	53.8	13	7.7
589 Lompoc	6	33.3	9	11.1
590 Fort Bragg	0	*	0	*
591 Whittier	49	61.2	59	42.4
592 Pittsburg	15	73.3	24	37.5
593 San Mateo	43	53.5	40	40.0

Table A (continued)

Number of Tests (*n*) and Test Fail Rates Over All Attempts on the English DL 5 (Rev. 6/05) for Original and Renewal Applicants in Each Field Office

Reporting unit number and field office	Originals		Renewals	
	n	Fail rate (%)	*n*	Fail rate (%)
594 Tulare	13	69.2	12	58.3
595 Lancaster	38	65.8	34	38.2
596 Oceanside	41	46.3	55	23.6
597 Brawley	6	100.0	4	25.0
598 Davis	9	55.6	11	18.2
599 Daly City	89	58.4	53	43.4
601 Paradise	2	100.0	9	33.3
602 Sacramento-South	66	71.2	47	53.2
603 Coalinga	1	0.0	0	*
604 Oakland Coliseum	48	68.8	26	69.2
605 Laguna Hills	0	*	0	*
606 Bellflower	57	71.9	49	46.9
607 Fullerton	0	*	1	100.0
608 Torrance	28	35.7	37	45.9
609 Hawthorne	42	64.3	18	38.9
610 Inglewood	32	65.6	67	61.2
611 Westminister	44	54.5	67	26.9
612 Rancho Cucamonga	49	59.2	50	42.0
613 Chula Vista	61	63.9	43	34.9
614 Spring Valley	0	*	0	*
615 Delano	4	75.0	9	77.8
616 Santa Monica	35	22.9	36	11.1
617 Lincoln Park	42	59.5	24	29.2
618 West Covina	96	67.7	79	45.6
619 San Pedro	24	70.8	28	39.3
620 Escondido	33	39.4	49	36.7
621 Fairfield	14	71.4	16	31.3
622 Lodi	21	52.4	29	48.3
623 Gilroy	8	75.0	10	40.0
624 Walnut Creek	15	26.7	53	32.1
625 Carmichael	34	47.1	78	34.6

Table A (continued)

Number of Tests (*n*) and Test Fail Rates Over All Attempts on the English DL 5 (Rev. 6/05) for Original and Renewal Applicants in Each Field Office

Reporting unit number and field office	Originals		Renewals	
	n	Fail rate (%)	*n*	Fail rate (%)
626 Redlands	27	48.1	39	43.6
627 Garberville	1	0.0	2	0.0
628 Costa Mesa	24	45.8	25	20.0
629 Victorville	42	69.0	35	34.3
630 Santa Paula	0	*	1	0.0
631 Pleasanton	27	44.4	32	25.0
632 Santa Clara	79	35.4	65	46.2
633 Reedley	22	77.3	10	30.0
634 Petaluma	17	47.1	29	41.4
635 Hemet	23	78.3	40	37.5
636 Oxnard	0	*	0	*
637 Winnetka	56	57.1	61	44.3
638 Twenty-nine Palms	23	60.9	8	50.0
639 Mount Shasta	4	0.0	3	66.7
640 Los Gatos	22	27.3	44	11.4
641 Banning	6	66.7	18	27.8
642 Tracy	7	71.4	16	31.3
643 Fall River Mills	0	*	2	100.0
644 Fremont	34	52.9	27	29.6
645 Orland	0	*	0	*
646 Fresno-North	28	71.4	32	43.8
647 King City	0	*	0	*
648 San Clemente	9	0.0	49	26.5
649 Turlock	14	57.1	30	26.7
650 Los Banos	11	54.5	10	50.0
655 Folsom	17	47.1	31	35.5
656 Riverside-East	58	69.0	33	42.4
657 Fontana	40	65.0	28	42.9
658 Manteca	8	50.0	15	26.7
659 Palm Springs	25	44.0	28	46.4
660 Shafter	8	37.5	10	60.0

Table A (continued)

Number of Tests (*n*) and Test Fail Rates Over All Attempts on the English DL 5 (Rev. 6/05) for Original and Renewal Applicants in Each Field Office

Reporting unit number and field office	Originals		Renewals	
	n	Fail rate (%)	*n*	Fail rate (%)
661 Arvin	7	85.7	8	62.5
662 Newhall	32	37.5	41	34.1
663 Thousand Oaks	13	7.7	35	20.0
668 Santa Teresa	23	47.8	26	42.3
669 El Cajon	0	*	1	100.0
670 Goleta	2	50.0	14	35.7
672 Temecula	37	43.2	47	27.7
673 Rocklin	8	62.5	21	42.9
676 Poway	20	55.5	32	25.0
677 San Ysidro	37	62.2	20	40.0
679 Bakersfield-Southwest	21	52.4	26	38.5
680 Simi Valley	6	50.0	23	30.4
686 Novato	3	33.3	11	9.1
687 Lake Isabella	3	66.7	3	33.3
690 Palmdale	23	73.9	12	33.3
696 Escondido	1	100.0	1	0.0
830 Clear Lake	2	50.0	6	33.3
859 Camp Pendleton	19	36.8	7	14.3

Note. Office fail rates that are based on fewer than 20 test forms are likely to be unreliable and should not be taken as accurate estimates. Fail rates are not presented for the English provisional or Spanish tests because too few of these tests were collected to compute accurate estimates for most of the offices. An asterisk (*) entry indicates that the fail rate could not be computed because no test forms of that type were received. King City and Oxnard were the only offices that did not send in any tests. Blythe and Colusa submitted only English DL 5T provisional tests and so their data are not included in the table. Data are also not shown for Alturas, Fort Bragg, and Laguna Hills because these offices did not submit any tests with the appropriate revision dates.

Appendix B

Item Statistics for Original Applicants on the English DL 5 First Attempt

Table B1

Percentage of Original Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5 First Attempt

Item	Answer choice	Form 1 (n = 222)	Form 2 (n = 225)	Form 3 (n = 213)	Form 4 (n = 217)	Form 5 (n = 206)	Form 6 (n = 200)	Form 7 (n = 225)	Form 8 (n = 200)	Form 9 (n = 218)	Form 10 (n = 213)
1	a	77.5	90.7	74.8	12.0	8.8	12.1	3.6	76.9	7.3	81.6
	b	6.8	3.1	10.0	86.1	77.1	83.9	5.4	8.5	89.0	8.0
	c	15.8	6.2	15.2	1.9	14.1	4.0	91.0	14.6	3.7	10.4
2	a	7.2	80.4	87.2	12.2	5.4	80.0	6.3	91.5	75.0	91.5
	b	6.8	8.4	9.0	75.6	3.9	6.0	79.4	4.5	7.9	3.3
	c	86.0	11.1	3.8	12.2	90.7	14.0	14.3	4.0	17.1	5.2
3	a	78.3	91.1	89.2	67.6	5.9	88.4	17.5	5.0	79.6	5.6
	b	18.1	4.0	6.6	20.8	11.8	4.5	79.8	10.1	**0.9**	6.6
	c	3.6	4.9	4.2	11.6	82.4	7.1	2.7	84.9	19.4	87.8
4	a	9.0	8.1	10.9	5.1	2.9	89.5	2.7	8.6	2.3	9.4
	b	83.8	78.5	**0.9**	2.3	80.0	**0.0**	2.7	15.2	88.5	71.8
	c	7.2	13.5	88.2	92.6	17.1	10.5	94.6	76.1	9.2	18.8
5	a	16.2	4.0	23.0	86.0	**1.5**	10.1	6.8	85.2	81.7	6.2
	b	5.9	91.5	75.1	**1.9**	**0.5**	13.1	73.0	7.7	11.9	88.6
	c	77.9	4.5	**1.9**	12.1	98.0	76.9	20.3	7.1	6.4	5.2
6	a	5.4	61.6	77.7	73.2	17.6	29.3	17.9	5.6	4.1	21.2
	b	9.5	28.6	16.1	14.1	10.7	64.6	75.8	3.0	90.8	72.2
	c	85.1	9.8	6.2	12.7	71.7	6.1	6.3	91.4	5.0	6.6
7	a	9.6	14.9	64.2	3.7	**2.0**	26.3	80.1	82.4	4.1	66.7
	b	83.1	12.6	27.8	2.8	97.1	68.2	7.7	11.1	89.4	14.1
	c	7.3	72.5	8.0	93.5	**1.0**	5.6	12.2	6.5	6.4	19.2
8	a	9.0	2.2	2.4	93.0	**0.0**	56.1	92.4	5.5	89.9	3.3
	b	90.1	3.1	79.1	4.7	4.9	21.4	6.3	83.4	**1.8**	79.2
	c	0.9	94.6	18.5	2.3	95.1	22.4	**1.3**	11.1	8.3	17.5
9	a	18.7	37.1	7.5	2.8	8.9	**1.5**	19.2	7.5	14.7	83.6
	b	65.8	59.4	17.0	90.7	83.7	23.2	22.3	76.4	61.8	**1.9**
	c	15.5	3.6	75.5	6.5	7.4	75.3	58.5	16.1	23.5	14.6
10	a	14.2	19.6	72.6	5.1	18.3	7.0	75.9	5.0	83.5	**1.4**
	b	70.3	9.8	24.1	74.3	65.8	6.5	11.2	95.0	3.2	6.1
	c	15.5	70.7	3.3	20.6	15.8	86.4	12.9	**0.0**	13.3	92.5
11	a	75.7	4.9	9.5	4.1	20.0	77.6	41.3	9.6	10.1	12.3
	b	12.4	88.4	3.8	81.6	72.7	14.3	**51.1**	82.3	85.3	81.6
	c	11.9	6.7	86.7	14.3	7.3	8.2	7.6	8.1	4.6	6.1
12	a	12.6	73.5	25.0	7.9	3.9	17.6	4.5	5.6	29.0	82.0
	b	63.5	14.8	61.1	59.7	95.1	4.0	89.7	82.1	56.2	16.1
	c	23.9	11.7	13.9	32.4	**1.0**	78.4	5.8	12.2	14.7	**1.9**
13	a	5.9	4.0	**0.0**	8.4	9.3	12.0	3.6	18.7	2.3	81.2
	b	86.5	14.7	**0.9**	82.3	11.3	3.0	85.1	66.2	85.7	13.1
	c	7.7	81.3	99.1	9.3	79.4	85.0	11.3	15.2	12.0	5.6
14	a	7.7	11.6	21.0	71.6	86.3	2.5	11.2	30.3	88.0	4.3
	b	12.3	4.9	6.2	16.7	13.2	4.5	77.1	63.1	5.1	93.4
	c	80.0	83.6	72.9	11.6	**0.5**	93.0	11.7	6.6	6.9	2.4
15	a	82.9	4.0	18.6	**1.4**	79.5	6.0	81.2	22.8	4.6	2.3
	b	11.7	44.6	60.5	97.2	11.2	84.9	14.3	71.6	85.3	**0.5**
	c	5.4	**51.3**	21.0	**1.4**	9.3	9.0	4.5	5.6	10.1	97.2
16	a	3.2	64.0	8.9	3.3	3.9	82.5	27.8	6.0	11.5	4.7
	b	5.9	26.1	85.4	82.2	11.8	9.5	**1.8**	29.1	82.9	92.0
	c	91.0	9.9	5.6	14.5	84.3	8.0	70.4	64.8	5.5	3.3

Table B1 (continued)

Percentage of Original Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5 First Attempt

Item	Answer choice	Form 1 (*n* = 222)	Form 2 (*n* = 225)	Form 3 (*n* = 213)	Form 4 (*n* = 217)	Form 5 (*n* = 206)	Form 6 (*n* = 200)	Form 7 (*n* = 225)	Form 8 (*n* = 200)	Form 9 (*n* = 218)	Form 10 (*n* = 213)
17	a	25.3	66.7	10.9	82.7	3.9	20.5	5.8	7.1	3.7	16.5
	b	74.7	4.5	75.4	5.6	36.0	15.5	65.6	30.6	96.3	80.2
	c	**0.0**	28.8	13.7	11.7	60.1	64.0	28.6	62.2	**0.0**	3.3
18	a	91.4	5.4	3.8	12.7	93.7	12.1	8.5	18.6	10.6	20.5
	b	5.0	87.9	9.5	73.2	2.9	11.1	83.5	77.9	60.1	4.8
	c	3.6	6.7	86.7	14.1	3.4	76.8	8.0	3.5	29.4	74.8
19	a	69.5	**0.9**	4.7	39.8	13.7	11.2	13.4	90.4	17.5	11.8
	b	13.6	84.3	89.2	2.8	10.8	70.6	8.0	6.1	6.5	79.7
	c	16.8	14.8	6.1	**57.4**	75.5	18.3	78.6	3.5	76.0	8.5
20	a	4.5	9.4	21.2	**1.4**	27.9	83.0	66.8	12.7	4.6	89.7
	b	18.1	4.9	77.8	96.3	14.2	2.5	8.5	79.2	6.9	9.9
	c	77.4	85.7	**0.9**	2.3	57.8	14.5	24.7	8.1	88.5	**0.5**
21	a	6.8	5.4	**1.9**	**1.8**	27.3	**1.0**	11.2	85.4	3.3	5.2
	b	2.3	7.6	11.3	**1.8**	7.8	4.5	10.3	7.0	82.3	92.5
	c	91.0	87.0	86.9	96.3	64.9	94.5	78.5	7.5	14.4	2.3
22	a	17.4	8.6	**0.5**	89.4	3.9	91.5	9.4	88.3	87.5	4.2
	b	7.3	5.0	47.4	**0.9**	5.4	6.5	3.1	6.6	7.9	91.1
	c	75.2	86.5	**52.1**	9.7	90.7	2.0	87.4	5.1	4.6	4.7
23	a	**1.8**	4.9	33.6	82.0	3.4	5.1	73.9	4.5	88.0	6.1
	b	95.5	3.6	9.0	12.9	93.1	19.7	18.9	85.4	5.1	7.0
	c	2.7	91.6	57.3	5.1	3.4	75.3	7.2	10.1	6.9	86.9
24	a	5.0	95.1	4.7	12.6	9.8	4.5	74.4	4.5	6.9	91.1
	b	85.5	3.1	19.3	75.8	86.8	7.1	4.5	84.9	91.3	2.8
	c	9.5	**1.8**	75.9	11.6	3.4	88.4	21.1	10.6	**1.8**	6.1
25	a	88.3	79.1	87.7	13.4	59.8	16.3	33.9	86.9	**1.8**	6.6
	b	4.1	13.3	5.2	72.2	8.3	2.6	3.1	6.5	95.4	81.0
	c	7.7	7.6	7.1	14.4	31.9	81.1	**62.9**	6.5	2.8	12.3
26	a	80.3	83.0	78.8	22.2	90.6	9.5	76.8	63.3	4.6	86.3
	b	6.3	12.1	15.1	69.0	6.9	78.9	12.9	15.3	66.8	9.9
	c	13.5	4.9	6.1	8.8	2.5	11.6	10.3	21.4	28.6	3.8
27	a	28.5	11.2	10.9	2.8	65.0	9.1	16.5	75.4	2.8	4.2
	b	62.9	82.1	13.3	19.1	13.3	11.6	82.6	17.6	**1.4**	8.5
	c	8.6	6.7	75.8	78.1	21.7	79.3	**0.9**	7.0	95.9	87.3
28	a	11.3	84.4	2.8	**0.0**	2.9	**44.2**	9.0	19.6	3.7	**0.9**
	b	6.8	11.6	4.3	**1.4**	**1.5**	48.2	13.5	6.0	3.2	93.4
	c	82.0	4.0	92.9	98.6	95.6	7.6	77.6	74.4	93.1	5.6
29	a	76.1	13.8	4.2	62.1	14.2	76.4	2.2	84.4	84.3	**1.9**
	b	16.7	2.2	87.3	14.5	80.9	15.6	84.8	10.6	11.1	9.5
	c	7.2	84.0	8.5	23.4	4.9	8.0	12.9	5.0	4.6	88.6
30	a	7.7	**0.4**	4.2	9.3	2.4	85.4	**1.3**	87.4	10.1	24.6
	b	78.8	6.7	10.3	84.7	3.9	11.6	**1.3**	5.5	4.6	69.7
	c	13.5	92.9	85.4	6.0	93.7	3.0	97.3	7.0	85.3	5.7
31	a	91.9	7.6	17.9	10.2	6.8	3.0	15.3	5.1	8.3	38.5
	b	3.2	81.3	57.5	6.0	1.0	91.4	9.0	4.0	77.8	49.3
	c	5.0	11.1	24.5	83.7	92.2	5.6	75.7	90.9	13.9	12.2
32	a	4.1	93.8	4.2	91.2	75.9	18.7	6.7	14.8	89.9	43.1
	b	5.0	5.3	91.5	6.0	7.4	14.1	89.7	13.3	4.6	3.3
	c	91.0	**0.9**	4.2	2.8	16.7	67.2	3.6	71.9	5.5	53.6

Table B1 (continued)

Percentage of Original Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5 First Attempt

Item	Answer choice	Form 1 (n = 222)	Form 2 (n = 225)	Form 3 (n = 213)	Form 4 (n = 217)	Form 5 (n = 206)	Form 6 (n = 200)	Form 7 (n = 225)	Form 8 (n = 200)	Form 9 (n = 218)	Form 10 (n = 213)
33	a	10.9	86.2	**1.9**	88.9	18.6	11.1	9.0	12.6	75.1	18.8
	b	13.2	8.9	33.5	8.3	67.6	81.8	85.7	5.5	4.1	26.0
	c	75.9	4.9	64.6	2.8	13.7	7.1	5.4	81.9	20.7	55.3
34	a	88.2	89.3	5.2	31.9	82.0	23.7	7.6	13.8	6.0	77.9
	b	5.4	7.1	81.2	7.4	3.4	3.5	7.6	16.3	15.2	9.9
	c	6.3	3.6	13.6	60.6	14.6	72.7	84.8	69.9	78.8	12.2
35	a	63.8	5.8	3.3	10.2	83.8	**1.5**	**1.3**	11.3	83.5	68.5
	b	32.6	82.2	90.6	70.7	12.7	12.6	94.6	73.8	4.1	6.6
	c	3.7	12.0	6.1	19.1	3.4	85.9	4.0	14.9	12.4	24.9
36	a	74.7	74.4	88.7	6.0	3.9	5.6	5.8	85.2	5.5	83.4
	b	15.7	20.6	9.9	81.9	25.0	52.5	4.5	8.7	2.3	3.3
	c	9.7	4.9	**1.4**	12.1	71.1	41.9	89.7	6.1	92.2	13.3

Note. The entry for n at the top of each column is the number of usable first-attempt tests that were used to compute the item choice response rates in the column. Underlining of a percentage indicates that the answer choice was the correct response according to the official answer key. Shading indicates that the item needs to be reviewed and possibly revised due to the item pass rate being too low or too high. A boldface percentage indicates that the distracter selection rate is too low or too high and therefore the item may need to be revised or replaced.

Table B2

Item-Total Correlation for Each Item on Each Form of the English DL 5 for Original Applicants on the First Attempt

Item	Form 1 (n = 222)	Form 2 (n = 225)	Form 3 (n = 213)	Form 4 (n = 217)	Form 5 (n = 206)	Form 6 (n = 200)	Form 7 (n = 225)	Form 8 (n = 200)	Form 9 (n = 218)	Form 10 (n = 213)
1	.22	.28	.29	.17	.15	.23	.16	.21	.06	.16
2	.12	.20	.15	.22	.08	.16	.17	.23	.15	.19
3	.15	.30	.39	.28	.39	.36	.34	.34	.15	.25
4	.38	.30	.09	.41	.37	.41	.33	.41	.42	.21
5	.24	.43	.24	.16	.15	.37	.47	.40	.44	.26
6	.36	.22	.38	.16	.28	.04	.31	.35	.17	.57
7	.18	.18	.25	.30	.22	.53	.33	.33	.36	.34
8	.23	.37	.25	.26	.29	.11	.39	.32	.39	.28
9	.17	.31	.17	.26	.13	.33	.25	.30	.24	.28
10	.35	.15	.04	.47	.29	.55	.28	.42	.25	.27
11	.32	.27	.30	.41	.14	.10	.25	.27	.24	.26
12	.42	.18	.30	.19	.33	.28	.36	.44	.27	.35
13	.35	.24	.22	.48	.21	.29	.34	.41	.37	.29
14	.31	.18	.26	.38	.09	.35	.37	.06	.33	.22
15	.48	.04	.44	.32	.42	.24	.32	.22	.38	.22
16	.16	.18	.39	.24	.31	.25	.16	.38	.17	.29
17	.23	.41	.16	.41	.19	.16	.26	.31	.22	.44
18	.21	.32	.32	.44	.16	.35	.24	.23	.40	.32
19	.25	.38	.36	.25	.28	.39	.25	.36	.13	.34
20	.36	.24	.30	.25	.20	.30	.31	.35	.33	.21
21	.35	.25	.41	.27	.28	.32	.40	.37	.41	.32
22	.08	.34	.42	.32	.23	.35	.42	.30	.41	.39
23	.41	.46	.25	.43	.34	.52	.40	.38	.36	.17
24	.29	.29	.16	.33	.39	.31	.28	.33	.39	.39
25	.46	.32	.33	.49	.28	.35	.22	.34	.28	.33
26	.28	.30	.08	.43	.36	.20	.33	.37	.01	.20
27	.04	.31	.29	.30	.43	.40	.15	.43	.10	.20
28	.37	.28	.23	.18	.30	.32	.33	.36	.22	.31
29	.56	.37	.21	.21	.20	.34	.36	.48	.44	.31
30	.32	.43	.28	.32	.12	.38	.30	.35	.40	.29
31	.48	.46	.29	.34	.31	.40	.47	.35	.36	.22
32	.32	.13	.31	.25	.20	.30	.12	.36	.20	.36
33	.20	.31	.26	.21	.32	.35	.37	.43	.39	.21
34	.31	.41	.41	.13	.24	.38	.37	.43	.21	.28
35	.35	.43	.40	.34	.21	.27	.25	.06	.37	.14
36	.34	.22	.20	.21	.31	.27	.35	.07	.19	.28

Note. The entry for n at the top of each column is the number of useable first-attempt tests that were used to compute the items total correlations in the column. Shading indicates that an item needs to be reviewed and possibly revised or replaced because the item-total correlation is negative or less than .10.

Table B3

Percentage of Original Applicants Who Would Have Passed on the First Attempt if Different Cut-Points (Number Wrong) Had Been Used for Each Form of the English DL 5

Number wrong	Form 1 (*n* = 222)	Form 2 (*n* = 225)	Form 3 (*n* = 213)	Form 4 (*n* = 217)	Form 5 (*n* = 206)	Form 6 (*n* = 200)	Form 7 (*n* = 225)	Form 8 (*n* = 200)	Form 9 (*n* = 218)	Form 10 (*n* = 213)	Total (*N* = 2,139)
0	4.1	3.6	2.8	2.3	2.4	3.5	3.1	2.5	5.5	5.2	3.5
1	9.0	7.6	7.0	9.2	8.3	5.5	6.2	9.5	12.8	8.9	8.4
2	15.3	11.6	17.8	14.3	16.6	13.5	14.2	16.0	24.3	15.0	15.9
3	26.1	21.8	19.7	21.7	24.9	20.5	20.9	23.0	33.0	31.0	24.3
4	33.8	30.7	24.9	31.8	31.2	27.5	24.9	33.0	41.7	38.5	31.8
5	40.5	42.7	31.5	42.4	40.5	35.5	36.4	40.0	51.8	46.5	40.8
6	51.4	51.1	40.8	51.2	52.2	46.5	43.1	48.0	61.0	53.1	49.9
7	59.9	61.3	50.2	59.4	59.5	52.0	51.1	58.0	69.3	60.6	58.2
8	67.6	67.1	58.7	68.2	65.4	60.5	60.0	66.5	76.1	69.0	65.9
9	73.4	78.2	69.0	73.7	75.6	69.0	65.3	71.0	81.2	73.7	73.1
10	77.5	82.2	73.7	76.5	82.9	72.5	73.8	75.0	85.3	77.5	77.7
11	82.4	86.2	80.3	80.6	86.8	78.0	81.3	81.0	88.1	81.2	82.6
12	86.0	88.0	85.4	83.9	88.8	82.0	84.4	84.5	90.4	84.5	85.8
13	90.1	91.6	87.8	88.0	93.2	86.5	87.1	88.5	93.1	87.3	89.3
14	91.0	93.3	89.7	89.9	95.6	88.5	88.4	91.5	94.0	93.4	91.5
15	92.3	95.1	93.0	92.6	96.6	89.5	90.7	93.5	96.3	95.3	93.5
16+	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. The entry for *n* at the top of each column is the number of usable first-attempt tests that were used to compute the percentages in the column. The shaded line highlights the pass rates at the current passing standard of six allowable errors.

Table B4

Summary of Problem Items on Each Form of the English DL 5 for Original Applicants on the First Attempt

Problem indicator	Form 1 items	Form 2 items	Form 3 items	Form 4 items	Form 5 items	Form 6 items	Form 7 items	Form 8 items	Form 9 items	Form 10 items
Item-total correlation too low or negative[a]	22, 27	15	4, 10, 26		2, 14	6		14, 35, 36	1, 26	
Pass rate too high[b]	23	24	13	15, 20, 21, 28	5, 7, 8, 12, 28		30	10	17, 25, 27	15
Pass rate too low[c]		15	22, 23, 31	12, 19	20, 25	28, 36	11, 25		12	31, 32, 33
Pass rate too high or low and item-total correlation too low or negative		15								
Distracter selected too often[d]		15	22	19		28	11			
Distracter selected too infrequently[e]	17, 23	24, 30, 32	13, 20, 21, 22, 33, 36	5, 15, 20, 21, 22, 28	5, 7, 8, 12, 14, 28	4, 9, 21, 35	8, 16, 27, 30, 35	10	3, 8, 17, 24, 25, 27	9, 10, 12, 15, 20, 28, 29

Note. The entries in the table are the test question numbers on the form. An item may be represented in more than one problem category.

[a]The item-total correlation was negative or less than .10. [b]More than 95% of the applicants answered the item correctly. [c]Less than 60% of the applicants answered the item correctly. [d]A distracter was chosen more, or almost, as often as the correct answer. [e]A distracter was selected by 2% or fewer of the applicants.

Appendix C

Item Statistics for Renewal Applicants on the English DL 5 First Attempt

Table C1

Percentage of Renewal Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5 First Attempt

Item	Answer choice	Form 1 (*n* = 311)	Form 2 (*n* = 315)	Form 3 (*n* = 322)	Form 4 (*n* = 319)	Form 5 (*n* = 339)	Form 6 (*n* = 325)	Form 7 (*n* = 321)	Form 8 (*n* = 331)	Form 9 (*n* = 313)	Form 10 (*n* = 324)
1	a	84.8	92.4	85.3	10.1	7.7	7.4	3.1	79.7	8.4	81.1
	b	3.9	2.9	6.0	87.1	80.7	90.5	5.0	7.3	90.0	6.2
	c	11.3	4.8	8.8	2.8	11.6	2.2	91.8	13.0	1.6	12.7
2	a	**1.6**	83.7	91.6	6.3	5.3	86.4	7.2	93.7	86.0	91.6
	b	5.5	4.2	5.0	79.4	**1.8**	4.0	81.9	**1.5**	4.2	5.9
	c	92.9	12.1	3.4	14.3	92.9	9.6	10.9	4.8	9.7	2.5
3	a	84.4	89.8	92.2	75.9	6.8	93.2	12.2	2.7	90.1	6.9
	b	12.0	6.4	3.1	15.4	15.1	**0.3**	85.6	10.1	**0.6**	5.6
	c	3.6	3.8	4.7	8.8	78.0	6.5	2.2	87.2	9.3	87.5
4	a	4.2	8.0	14.6	**0.9**	6.5	89.2	2.2	5.2	2.6	9.4
	b	91.5	77.2	**0.6**	**1.6**	79.5	**0.9**	2.5	21.0	92.3	76.3
	c	4.2	14.7	84.7	97.5	14.0	9.8	95.3	73.9	5.1	14.4
5	a	5.5	2.3	17.6	88.1	3.3	13.6	6.6	84.1	89.0	**1.5**
	b	**1.6**	94.9	78.6	**1.9**	2.1	5.2	84.7	9.5	4.5	95.4
	c	92.9	2.9	3.8	10.1	94.7	81.2	8.8	6.4	6.5	3.1
6	a	**1.9**	63.3	76.2	81.1	12.8	18.3	27.0	8.2	2.9	19.3
	b	3.9	29.5	20.7	6.6	9.3	79.5	65.7	**1.8**	93.6	74.5
	c	94.2	7.1	3.1	12.3	77.9	2.2	7.2	90.0	3.5	6.2
7	a	8.1	12.6	76.3	**1.9**	2.7	15.8	88.1	88.4	2.9	69.7
	b	85.4	8.4	18.8	2.8	94.7	75.2	8.5	4.6	92.9	12.1
	c	6.5	79.0	5.0	95.3	2.7	9.0	3.4	7.0	4.2	18.3
8	a	6.1	5.8	6.9	92.8	**0.6**	75.6	90.9	2.8	96.2	**1.5**
	b	91.9	2.6	71.8	3.8	**1.8**	11.7	6.6	87.2	**1.6**	90.7
	c	1.9	91.7	21.3	3.5	97.6	12.7	2.5	10.1	2.2	7.7
9	a	8.5	39.9	11.9	2.5	6.9	**1.5**	20.8	3.6	10.7	84.2
	b	80.8	**56.3**	10.6	93.7	83.8	16.4	16.0	83.3	71.4	3.1
	c	10.7	3.9	77.5	3.8	9.3	82.1	63.2	13.0	17.9	12.7
10	a	7.1	17.5	76.8	4.1	5.0	3.1	77.2	2.4	91.4	**1.6**
	b	78.6	8.4	17.9	83.3	86.9	7.4	7.8	95.8	1.6	4.7
	c	14.2	74.1	5.3	12.6	8.0	89.5	15.0	**1.8**	7.0	93.8
11	a	81.4	**1.6**	10.6	**1.9**	25.8	72.1	43.6	9.8	2.6	8.0
	b	9.1	92.6	4.0	89.3	68.5	19.8	**54.5**	86.3	96.2	87.3
	c	9.4	5.8	85.4	8.8	5.7	8.0	**1.9**	4.0	**1.3**	4.6
12	a	7.1	70.8	16.6	2.2	2.7	11.1	**0.9**	5.5	28.6	78.3
	b	71.1	18.9	74.6	64.0	96.2	**0.9**	97.8	84.5	64.3	18.3
	c	21.8	10.3	8.8	33.8	**1.2**	88.0	1.2	10.0	7.1	3.4
13	a	4.9	2.2	**0.6**	6.4	4.1	6.5	3.8	14.1	**0.6**	83.8
	b	90.6	12.7	2.8	84.0	6.5	**1.6**	88.1	81.3	91.9	10.6
	c	4.5	85.0	96.6	9.6	89.3	91.9	8.2	4.6	7.4	5.6
14	a	2.9	14.7	18.7	83.3	90.5	**1.2**	14.5	22.2	88.7	7.1
	b	7.4	2.6	5.0	11.4	7.1	4.6	74.8	74.8	2.6	89.8
	c	89.7	82.7	76.3	5.4	2.4	94.2	10.7	3.0	8.7	3.1
15	a	92.3	2.3	8.6	**0.6**	78.0	3.1	78.4	16.8	5.5	**1.2**
	b	3.2	49.2	78.1	98.1	6.5	91.0	18.2	80.8	86.2	**0.6**
	c	4.5	**48.6**	13.3	1.3	15.4	5.9	3.4	2.4	8.4	98.1
16	a	8.7	63.3	4.7	2.2	2.1	84.9	25.5	6.4	2.9	8.0
	b	4.5	27.0	92.5	86.1	3.8	5.8	2.5	21.5	92.9	88.2
	c	86.7	9.6	2.8	11.7	94.1	9.2	72.0	72.1	4.2	3.7

Table C1 (continued)

Percentage of Renewal Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5 First Attempt

Item	Answer choice	Form 1 (n = 311)	Form 2 (n = 315)	Form 3 (n = 322)	Form 4 (n = 319)	Form 5 (n = 339)	Form 6 (n = 325)	Form 7 (n = 321)	Form 8 (n = 331)	Form 9 (n = 313)	Form 10 (n = 324)
17	a	13.2	76.0	11.5	87.8	**1.2**	17.0	5.7	9.2	3.2	18.8
	b	85.8	5.8	71.4	3.8	28.5	15.5	75.9	13.8	96.8	77.2
	c	**1.0**	18.2	17.1	8.3	70.3	67.5	18.4	77.1	**0.0**	4.0
18	a	95.8	6.0	3.4	14.2	95.9	13.6	13.5	14.0	4.5	10.8
	b	3.2	89.8	6.0	72.9	**1.8**	10.2	80.8	81.5	74.3	**1.2**
	c	**1.0**	4.1	90.6	12.9	2.4	76.2	5.7	4.6	21.2	88.0

Note. The entry for *n* at the top of each column is the number of usable first-attempt tests that were used to compute the item choice response rates in the column. Underlining of a percentage indicates that the answer choice was the correct response according to the official answer key. Shading indicates that the item needs to be reviewed and possibly revised due to the item pass rate being too low or too high. A boldface percentage indicates that the distracter selection rate is too low or too high.

Table C2

Item-Total Correlation for Each Item on Each Form of the English DL 5 for Renewal Applicants on the First Attempt

Item	Form 1 (n = 311)	Form 2 (n = 315)	Form 3 (n = 322)	Form 4 (n = 319)	Form 5 (n = 339)	Form 6 (n = 325)	Form 7 (n = 321)	Form 8 (n = 331)	Form 9 (n = 313)	Form 10 (n = 324)
1	.20	.11	.24	.42	.17	.12	.07	.17	.13	.14
2	.20	.28	.38	.17	.21	.11	.14	.21	.16	.19
3	.26	.29	.48	.23	.34	.23	.35	.27	.11	.09
4	.22	.27	.13	.42	.12	.19	.31	.25	.37	.01
5	.29	.25	.29	.39	.07	.13	.28	.21	.34	.33
6	.36	.17	.18	.15	.19	.13	.03	.33	.15	.27
7	.15	.20	.24	.55	.20	.31	.19	.26	.25	.34
8	.32	.27	.28	.29	.21	.24	.29	.27	.15	.24
9	.22	.19	.23	.42	.06	.30	.16	.26	.23	.32
10	.20	.09	.25	.36	.24	.22	.21	.20	.27	.19
11	.21	.24	.38	.30	.15	.10	.28	.23	.14	.23
12	.22	.29	.31	.17	.38	.25	.16	.22	.25	.44
13	.45	.23	.38	.40	.11	.29	.13	.48	.33	.30
14	.43	.21	.36	.32	.09	.25	.38	.25	.27	.37
15	.36	.08	.16	.46	.21	.08	.19	.08	.24	.30
16	.32	.05	.39	.20	.09	.20	.15	.30	.24	.24
17	.25	.23	.28	.41	.15	.16	.25	.29	.19	.17
18	.24	.11	.38	.25	.04	.26	.28	.33	.22	.19

Note. The entry for n at the top of each column is the number of usable first-attempt tests that were used to compute the item-total correlations in the column. Shading indicates that an item needs to be reviewed and possibly revised or replaced because the item-total correlation is negative or less than .10.

Table C3

Percentage of Renewal Applicants Who Would Have Passed on the First Attempt if Different Cut-Points (Number Wrong) Had Been Used for Each Form of the English DL 5

Number wrong	Form 1 (n = 311)	Form 2 (n = 315)	Form 3 (n = 322)	Form 4 (n = 319)	Form 5 (n = 339)	Form 6 (n = 325)	Form 7 (n = 321)	Form 8 (n = 331)	Form 9 (n = 313)	Form 10 (n = 324)	Total (N = 3,220)
0	21.5	1.9	8.1	15.7	13.0	11.4	5.0	12.4	18.2	13.0	12.0
1	43.4	9.2	26.7	37.6	32.2	28.9	21.2	32.3	43.1	35.5	31.0
2	60.5	25.4	46.6	56.4	56.3	48.3	34.6	47.4	64.5	57.1	49.7
3	75.9	45.1	63.0	72.4	72.6	66.8	53.0	64.7	79.9	70.4	66.4
4	84.9	62.2	74.2	82.8	83.2	77.8	66.0	76.1	89.5	81.2	77.8
5	91.0	75.2	83.5	89.7	90.9	86.2	76.6	84.3	93.3	88.9	86.0
6	94.2	85.7	90.1	93.1	96.5	91.7	86.0	90.0	94.9	92.3	91.5
7	96.1	90.8	92.5	94.7	98.5	96.0	92.2	93.7	96.5	96.0	94.7
8	97.4	96.2	95.3	95.3	99.1	97.5	95.0	95.8	97.4	97.5	96.7
9	98.1	97.5	96.0	97.2	99.4	99.1	97.5	97.0	98.7	98.8	97.9
10+	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. The entry for n at the top of each column is the number of usable first-attempt tests that were used to compute the percentages in the column. The shaded line highlights the pass rates at the current passing standard of three allowable errors.

Table C4

Summary of Problem Items on Each Form of the English DL 5 for Renewal Applicants on the First Attempt

Problem indicator	Form 1 items	Form 2 items	Form 3 items	Form 4 items	Form 5 items	Form 6 items	Form 7 items	Form 8 items	Form 9 items	Form 10 items
Item-total correlation too low or negative[a]		10, 15, 16			5, 9, 14, 16, 18	15	1, 6	15		3, 4
Pass rate too high[b]	18		13	4, 7	8, 18		4, 12	10	8, 11	5, 15
Pass rate too low[c]		15								
Pass rate too high or low and item-total correlation too low or negative		15					11			
Distracter selected too often[d]		9, 15					11			
Distracter selected too infrequently[e]	2, 5, 16, 17, 18	11	4, 13	4, 5, 7, 11, 15	2, 8, 12, 17, 18	3, 4, 9, 12, 13, 14	11, 12	2, 6, 10	3, 8, 11, 13, 17	5, 8, 10, 15, 18

Note. The entries in the table are the test question numbers on the form. An item may be represented in more than one problem category.

[a]The item-total correlation was negative or less than .10. [b]More than 95% of the applicants answered the item correctly. [c]Less than 60% of the applicants answered the item correctly. [d]A distracter was chosen more, or almost, as often as the correct answer. [e]A distracter was selected by 2% or fewer of the applicants.

Appendix D

Item Statistics for Provisional Applicants on the English DL 5T First Attempt

Table D1

Percentage of Provisional Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5T First Attempt

Item	Answer choice	Form 1 (n = 247)	Form 2 (n = 255)	Form 3 (n = 251)	Form 4 (n = 247)	Form 5 (n = 239)
1	a	99.2	90.1	28.6	96.4	5.0
	b	**0.8**	3.2	56.7	2.4	12.6
	c	**0.0**	6.7	14.7	**1.2**	82.4
2	a	2.4	33.9	10.0	87.0	2.1
	b	82.6	63.0	87.6	8.9	95.8
	c	15.0	3.1	2.4	4.0	2.1
3	a	89.0	14.2	33.9	67.6	69.9
	b	6.1	73.1	19.5	10.9	6.8
	c	4.9	12.6	46.6	21.5	23.3
4	a	76.9	90.1	92.8	21.9	19.3
	b	21.5	5.5	6.0	70.0	2.1
	c	**1.6**	4.3	**1.2**	8.1	78.6
5	a	4.5	**2.0**	18.7	6.5	92.1
	b	4.9	52.2	80.1	87.0	2.5
	c	90.7	**45.8**	**1.2**	6.5	5.4
6	a	98.4	**1.2**	39.2	78.1	10.9
	b	**0.4**	97.6	50.4	15.0	79.4
	c	**1.2**	**1.2**	10.4	6.9	9.7
7	a	80.9	89.4	20.7	3.2	38.4
	b	8.9	**1.6**	**1.2**	6.5	3.0
	c	10.2	9.1	78.1	90.3	58.6
8	a	11.0	5.5	46.2	12.1	11.3
	b	8.6	**0.4**	**47.8**	83.0	82.8
	c	80.4	94.1	6.0	4.9	5.9
9	a	2.8	82.3	4.5	15.1	96.6
	b	92.7	11.4	22.3	71.8	**1.7**
	c	4.5	6.3	73.3	13.1	**1.7**
10	a	3.6	**1.2**	78.0	19.4	3.4
	b	94.3	81.0	13.2	12.1	4.2
	c	2.0	17.8	8.8	68.4	92.4
11	a	92.3	5.9	2.4	93.1	2.9
	b	3.3	87.4	16.3	2.8	3.4
	c	4.5	6.7	81.3	4.0	93.7
12	a	14.6	85.0	40.3	**1.2**	19.3
	b	14.6	5.9	52.8	96.0	24.8
	c	70.7	9.1	6.9	2.8	55.9
13	a	6.1	64.4	**0.8**	5.3	82.8
	b	3.7	28.9	2.4	11.8	11.3
	c	90.2	6.7	96.8	82.9	5.9
14	a	15.9	84.5	10.8	14.6	**1.3**
	b	73.2	12.7	83.7	80.1	83.3
	c	11.0	2.8	5.6	5.3	15.5
15	a	**0.8**	5.1	**1.2**	16.6	17.7
	b	4.1	81.9	91.2	2.8	78.5
	c	95.1	13.0	7.6	80.6	3.8
16	a	9.4	8.3	87.6	6.5	83.7
	b	80.8	74.7	**0.8**	93.5	8.8
	c	9.8	17.0	11.6	**0.0**	7.5

Table D1 (continued)

Percentage of Provisional Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5T First Attempt

Item	Answer choice	Form 1 (n = 247)	Form 2 (n = 255)	Form 3 (n = 251)	Form 4 (n = 247)	Form 5 (n = 239)
17	a	2.8	**0.8**	5.2	84.2	11.8
	b	17.9	98.0	69.2	14.2	76.1
	c	79.3	**1.2**	25.6	**1.6**	12.2
18	a	16.2	**0.4**	96.0	12.2	87.8
	b	83.8	3.2	3.6	5.7	4.2
	c	**0.0**	96.4	**0.4**	82.1	8.0
19	a	23.6	4.7	5.2	76.3	75.1
	b	71.5	4.3	5.6	14.7	4.6
	c	4.9	90.9	89.2	9.0	20.3
20	a	74.5	**2.0**	7.6	6.1	2.1
	b	18.2	68.5	5.6	20.3	92.9
	c	7.3	29.5	86.9	73.6	5.0
21	a	15.1	42.5	12.0	**1.2**	84.9
	b	77.1	4.3	76.8	95.5	10.0
	c	7.8	**53.1**	11.2	3.2	5.0
22	a	15.0	96.1	**1.2**	7.3	7.5
	b	8.9	**0.8**	2.4	89.1	7.9
	c	76.1	3.1	96.4	3.6	84.5
23	a	90.7	3.2	27.6	6.9	69.3
	b	6.9	16.2	8.4	83.0	5.9
	c	2.4	80.6	64.0	10.1	24.8
24	a	4.0	**2.0**	8.0	87.4	16.4
	b	15.0	26.5	**0.4**	**0.8**	71.0
	c	81.0	71.5	91.6	11.8	12.6
25	a	76.1	**1.6**	66.5	7.3	2.9
	b	6.1	96.9	12.7	14.6	91.2
	c	17.8	**1.6**	20.7	78.1	5.9
26	a	17.8	22.8	14.3	2.8	20.6
	b	72.7	2.8	20.3	82.5	74.8
	c	9.5	74.4	65.3	14.6	4.6
27	a	68.0	20.3	83.5	91.1	**1.7**
	b	23.5	71.7	8.5	2.4	**1.3**
	c	8.5	8.0	8.1	6.5	97.1
28	a	97.6	**0.4**	3.6	10.2	2.1
	b	**1.2**	97.6	4.0	11.4	6.3
	c	**1.2**	**2.0**	92.4	78.4	91.6
29	a	**1.2**	90.6	3.2	93.5	4.2
	b	6.1	5.9	5.2	4.9	4.6
	c	92.7	3.5	91.6	**1.6**	91.2
30	a	**1.6**	6.7	76.5	88.2	86.0
	b	95.5	87.0	9.6	7.3	3.0
	c	2.8	6.3	13.9	4.5	11.0
31	a	**0.8**	2.4	3.2	94.7	83.3
	b	90.7	10.8	93.6	5.3	12.6
	c	8.5	86.8	3.2	**0.0**	4.2
32	a	3.7	88.5	8.8	**1.6**	7.1
	b	64.9	**0.8**	88.8	3.2	90.4
	c	31.4	10.7	2.4	95.1	2.5

Table D1 (continued)

Percentage of Provisional Applicants Selecting Each Answer Choice for Each Item on Each Form of the English DL 5T First Attempt

Item	Answer choice	Form 1 (n = 247)	Form 2 (n = 255)	Form 3 (n = 251)	Form 4 (n = 247)	Form 5 (n = 239)
33	a	**1.2**	8.3	71.8	2.8	85.8
	b	89.5	86.6	10.1	89.9	4.2
	c	9.3	5.1	18.1	7.3	10.0
34	a	13.4	**0.8**	8.8	58.9	**0.8**
	b	4.9	**0.0**	**1.6**	30.1	97.1
	c	81.8	99.2	89.6	11.0	2.1
35	a	83.7	2.8	2.8	13.4	5.0
	b	2.4	4.7	**0.4**	16.2	81.9
	c	13.8	92.5	96.8	70.4	13.0
36	a	16.3	3.1	79.7	86.9	86.9
	b	12.2	**2.0**	14.3	7.8	10.5
	c	71.5	94.9	6.0	5.3	2.5
37	a	8.9	93.7	20.1	**1.6**	3.8
	b	**2.0**	2.8	8.4	5.3	94.1
	c	89.1	3.6	71.5	93.1	2.1
38	a	11.4	69.0	39.0	2.4	9.3
	b	85.8	7.9	58.6	84.2	78.5
	c	2.8	23.0	2.4	13.4	12.2
39	a	**2.0**	59.4	10.5	91.1	21.1
	b	91.9	16.1	10.9	6.5	76.4
	c	6.1	24.4	78.6	2.4	2.5
40	a	**2.0**	3.9	**0.0**	5.3	86.1
	b	89.1	86.6	92.0	82.0	8.0
	c	8.9	9.4	8.0	12.7	5.9
41	a	71.0	**2.0**	4.4	81.3	16.4
	b	19.6	2.4	2.4	13.8	20.2
	c	9.4	95.7	93.2	4.9	63.4
42	a	6.5	9.8	8.1	4.5	73.4
	b	77.7	86.2	15.0	90.7	11.4
	c	15.8	3.9	76.8	4.9	15.2
43	a	**0.0**	3.2	23.2	11.8	7.2
	b	**0.8**	17.2	3.2	83.3	86.9
	c	99.2	79.6	73.6	4.9	5.9
44	a	**0.4**	16.5	84.4	3.2	3.4
	b	**2.0**	81.9	2.8	81.0	90.8
	c	97.6	**1.6**	12.8	15.8	5.9
45	a	5.7	93.3	85.6	95.5	4.6
	b	80.7	**0.8**	7.6	**0.4**	93.7
	c	13.5	5.9	6.8	4.0	**1.7**
46	a	96.4	4.0	96.0	85.4	84.5
	b	2.4	90.9	3.6	8.5	12.6
	c	**1.2**	5.1	**0.4**	6.1	2.9

Note. The entry for n at the top of each column is the number of usable first-attempt tests that were used to compute the item choice response rates in the column. Underlining of a percentage indicates that the answer choice was the correct response according to the official answer key. Shading indicates that the item needs to be reviewed and possibly revised due to the item pass rate being too low or too high. A boldface percentage indicates that the distracter selection rate is too low or too high.

Table D2

Item-Total Correlation for Each Item on Each Form of the English DL 5T for Provisional Applicants on the First Attempt

Item	Form 1 (n = 247)	Form 2 (n = 255)	Form 3 (n = 251)	Form 4 (n = 247)	Form 5 (n = 239)
1	.06	.30	.21	.19	.12
2	.16	.30	.13	.23	.10
3	.24	.43	.24	.36	.30
4	.20	.23	.20	.10	.23
5	.25	.37	.31	.14	.20
6	.01	.24	.27	.19	.09
7	.36	.12	.32	.28	.35
8	.29	.29	.26	.21	.27
9	.26	.26	.41	.33	.18
10	.30	.40	.17	.45	.28
11	.22	.14	.37	.10	.27
12	.08	.24	.25	.32	.23
13	.18	.16	.26	.22	.33
14	.33	.21	.27	.26	.39
15	.27	.29	.30	.10	.38
16	.13	.38	.20	.26	.25
17	.18	.18	.26	.21	.30
18	.29	.20	.19	.16	.28
19	.25	.17	.30	.33	.29
20	-.01	.29	.33	.30	.28
21	.20	.33	.25	.21	.25
22	.11	.22	.30	.28	.18
23	.24	.16	.30	.09	.28
24	.27	.35	.30	.21	.35
25	.46	.20	.25	.31	.23
26	.26	.17	.13	.32	.23
27	.18	.27	.18	.09	.31
28	.12	.15	.32	.24	.24
29	.17	.18	.17	.25	.46
30	.07	.22	.37	.29	.28
31	.11	.09	.15	.17	.25
32	.37	.24	.24	.20	.35
33	.23	.20	.37	.14	.45
34	.14	.07	.24	.28	.27
35	.20	.10	.31	.17	.12
36	.15	.21	.36	.17	.39
37	.35	.24	.25	.28	.30
38	.30	.23	.11	.31	.26
39	.21	.46	.32	.30	.27
40	.19	.23	.12	.14	.31
41	.35	.28	.40	.20	.30
42	.29	.16	.37	.18	.32
43	-.02	-.09	.27	.19	.33
44	.14	.18	.23	.24	.24
45	.25	.14	.20	.11	.23
46	.14	.22	.13	.13	.11

Note. The entry for n at the top of each column is the number of usable first-attempt tests that were used to compute the item-total correlations in the column. Shading indicates that an item needs to be reviewed and possibly revised or replaced because the item-total correlation is negative or less than .10.

Table D3

Percentage of Provisional Applicants Who Would Have Passed on the First Attempt if Different Cut-Points (Number Wrong) Had Been Used for Each Form of the English DL 5T

Number Missed	Form 1 (n = 247)	Form 2 (n = 255)	Form 3 (n = 251)	Form 4 (n = 247)	Form 5 (n = 239)	Total (N = 1,240)
0	4.0	4.3	2.0	3.2	2.1	2.3
1	10.1	9.8	4.4	10.5	7.1	7.3
2	16.2	14.6	7.6	13.8	13.4	12.1
3	22.7	20.5	11.2	21.1	22.2	18.3
4	29.6	29.5	19.9	30.0	31.4	26.2
5	38.9	40.6	24.3	42.1	41.8	35.2
6	47.4	47.2	29.9	48.6	49.0	43.0
7	56.7	53.9	37.5	56.3	53.6	50.2
8	62.8	61.4	45.4	62.3	62.3	57.3
9	70.9	68.9	53.0	68.8	69.0	64.5
10	78.5	75.2	62.9	75.3	74.9	72.1
11	82.6	81.5	70.1	81.4	78.2	77.5
12	86.2	87.4	73.7	87.9	81.6	82.1
13	89.5	89.0	80.5	91.1	86.6	87.0
14	92.7	90.9	82.5	93.5	90.0	89.5
15	94.7	92.5	86.1	95.1	92.1	91.8
16	95.5	94.9	88.0	96.4	93.3	93.1
17	97.6	96.9	90.8	96.8	94.1	94.8
18	99.2	97.2	93.2	97.6	95.8	96.5
19	99.6	98.8	94.8	98.8	96.2	97.3
20+	100.0	100.0	100.0	100.0	100.0	100.0

Note. The entry for n at the top of each column is the number of usable first-attempt tests that were used to estimate the percentages in the column. The shaded line highlights the pass rates at the current passing standard of three allowable errors.

Table D4

Summary of Problem Items on Each Form of the English DL 5T for Provisional Applicants on the First Attempt

Problem indicator	Form 1 items	Form 2 items	Form 3 items	Form 4 items	Form 5 items
Item-total correlation too low or negative[a]	1, 6, 12, 20, 30, 43	31, 34, 43		23, 27	6
Pass rate too high[b]	1, 6, 15, 28, 30, 43, 44, 46	6, 17, 25, 28, 34, 41	13, 18, 46	1, 12, 21, 32, 45	2, 9, 34
Pass rate too low[c]		5, 21, 39	1, 3, 6, 8, 12, 38	34	7, 12
Pass rate too high or low and item-total correlation too low or negative	1, 6, 30	34			
Distracter selected too often[d]		5, 21	8, 12		
Distracter selected too infrequently[e]	1, 4, 6, 15, 18, 28, 30, 31, 33, 37, 39, 40, 43, 44, 46	5, 6, 7, 8, 10, 17, 18, 20, 22, 24, 25, 28, 32, 34, 36, 41, 44, 45	4, 5, 7, 13, 15, 16, 18, 21, 24, 29, 31, 32, 37, 46	1, 12, 16, 17, 21, 24, 29, 31, 32, 37, 45	9, 14, 27, 34, 35

Note. The entries in the table are the test question numbers on the form. An item may be represented in more than one problem category.

[a]The item-total correlation was negative or less than .10. [b]More than 95% of the applicants answered the item correctly. [c]Less than 60% of the applicants answered the item correctly. [d]A distracter was chosen more, or almost, as often as the correct answer. [e]A distracter was selected by 2% or fewer of the applicants.